

TRABAJO ESPECIAL DE GRADO

**DESARROLLO DE UNA HERRAMIENTA DE ANÁLISIS DE DATOS DE
PRODUCCIÓN Y OPERATIVIDAD DE UN GRUPO DE POZOS, USANDO
MÉTODOS DE *DATA SCIENCE* Y *MACHINE LEARNING***

Presentado ante la Ilustre
Universidad Central de Venezuela
Por los Brs. Mijares M. Adriana A.
Sánchez L. John D.
Para optar al título de
Ingeniero de Petróleo

Caracas, 2020

TRABAJO ESPECIAL DE GRADO

**DESARROLLO DE UNA HERRAMIENTA DE ANÁLISIS DE DATOS DE
PRODUCCIÓN Y OPERATIVIDAD DE UN GRUPO DE POZOS, USANDO
MÉTODOS DE *DATA SCIENCE* Y *MACHINE LEARNING***

TUTOR ACADÉMICO:

Ing. Eliana Alvarado

Presentado ante la Ilustre
Universidad Central de Venezuela
Por los Brs. Mijares M. Adriana A.
Sánchez L. John D.
Para optar al título de
Ingeniero de Petróleo

Caracas, 2020

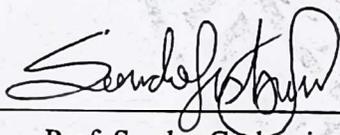
Caracas, 17 de marzo de 2021

FACULTAD DE INGENIERIA

Los abajo firmantes, miembros del jurado designado por el Consejo de Escuela de Ingeniería de Petróleo, para evaluar el Trabajo Especial de Grado presentado por los bachilleres Mijares Morales Adriana Alejandra y Sánchez León John David, titulado:

DESARROLLO DE UNA HERRAMIENTA DE ANÁLISIS DE DATOS DE PRODUCCIÓN Y OPERATIVIDAD DE UN GRUPO DE POZOS, USANDO MÉTODOS DE *DATA SCIENCE* Y *MACHINE LEARNING*

Consideran que el mismo cumple con los requisitos exigidos por el plan de estudios conducente al Título de Ingeniero de Petróleo y sin que ello signifique que se hacen solidarios con las ideas expuestas por los autores, lo declaran **APROBADO**.



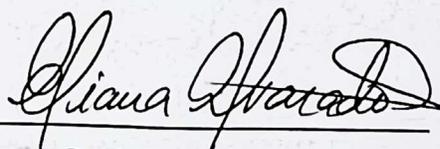
Prof. Sandro Gasbarri

Jurado



Prof. Adriana Zambrano

Jurado



Prof. Eliana Alvarado

Tutora

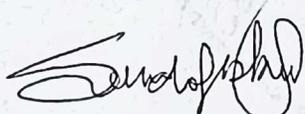
FACULTAD DE INGENIERIA

ACTA

Quienes suscriben, integrantes de la totalidad del jurado examinador del Trabajo Especial de Grado, titulado: “**DESARROLLO DE UNA HERRAMIENTA DE ANÁLISIS DE DATOS DE PRODUCCIÓN Y OPERATIVIDAD DE UN GRUPO DE POZOS, USANDO MÉTODOS DE DATA SCIENCE Y MACHINE LEARNING**”, presentado por los Bachilleres **ADRIANA ALEJANDRA MIJARES MORALES** y **JOHN DAVID SANCHEZ LEON**, titulares de la Cédula de Identidad No. 20.327.160 y 18.819.099 respectivamente. Para optar al título de Ingeniero de Petróleo, acordaron por unanimidad solicitarles a las autoridades correspondientes de la Facultad de Ingeniería, tengan a bien otorgar **MENCIÓN HONORÍFICA** al Trabajo de Grado antes mencionado.

Esta solicitud se hace con base a la originalidad del enfoque utilizado para tratar el tema presentado y a la excelencia, tanto de su contenido como de su presentación formal, atributos que lo convierten en una fuente de referencia útil.

Se levanta la presente Acta, en Caracas a los diez y siete (17) días del mes de marzo del año dos mil veintiuno.



Prof. Sandro Gasbarri
Jurado Principal



Prof. Adriana Zambrano
Jurado Principal



Prof. Eliana Alvarado
Tutor Académico

DEDICATORIA

“Todos los triunfos nacen cuando nos atrevemos a brillar”

A mi madre Lineke Morales, por su amor y esfuerzo. Por apoyarme siempre en cada uno de mis pasos y respetar mis decisiones. Por perdonarme y recibirme con los brazos abiertos cada vez que cometí un error.

A mi familia, por los buenos momentos, por ser únicos e incondicionales, por mi educación y valores que son su reflejo. Por enseñarme a ser leal y darme la oportunidad de crecer a su lado. Por enseñarme a soñar y luchar por mis sueños.

A ti, Héctor, por estar a mi lado siempre, apoyándome en las buenas y en las malas. Por tu amor incondicional y por mantenerme de pie, enfocada y nunca dejarme caer.

Por esto y mil razones más...

Quiero dedicarles este trabajo

Adriana Mijares

*A mis padres: María León, Marcelo Sánchez y Rosa Bohórquez.
A mi amiga y compañera en estos últimos años: Adriana Mijares.*

John Sánchez

AGRADECIMIENTOS

El camino ha sido largo, lleno de muchas experiencias y compañeros, muchos se han ido y otros, aún, siguen a nuestro lado, pero, sin importar donde estén, tengo mucho que agradecerles. Por los momentos vividos, por las lecciones aprendidas, por el apoyo, por las sonrisas, por las tristezas y por lo que soy hoy en día.

A Dios por ser mi guía espiritual, por acompañarme en silencio, y por darme la oportunidad de estar aquí.

A la Universidad Central de Venezuela, a la Facultad de Ingeniería y a la escuela de petróleo, por abrirme las puertas, por la formación recibida y por ser mi segundo hogar, dejándome experiencias y personas inolvidables.

A mi tutora Eliana Alvarado por darme la oportunidad de realizar este trabajo, por su valioso tiempo y por compartir su experiencia y conocimiento.

Al Sr. Raúl Páez y a la empresa Equinor Venezuela por darme las herramientas y los datos necesarios para desarrollar esta investigación.

A nuestros mentores y colaboradores, el Ing. Diego Manfre, Ing. Alejandro Primera, Ing. Carlos y al Ing. Osmer Parabavire, por compartir sus experiencias e introducirnos en un mundo tan innovador.

A mi familia y a mi novio, por ser parte de este logro, por las horas de traspaso, por su apoyo y amor incondicional.

A mis profesores Pedro Díaz, Neyfrank Ortuño, Yujeisly Reina, Miguel Castillejo, por los momentos compartidos, por su paciencia y por su amistad.

Al resto de mis profesores por su dedicación en mi formación.

A mis amigos y compañeros de clase que hicieron este camino mucho más fácil. Por los recuerdos, por las risas, por los momentos compartidos, por las horas de estudio, por ser parte de mi equipo y creer siempre en mí.

Gracias infinitas a todos los que me acompañaron durante este camino.

Adriana Mijares

Primeramente, agradezco a Dios por darme la oportunidad de culminar este proyecto, que representa los años de aprendizaje y esfuerzo a lo largo de la carrera. Le doy las gracias, además, por darme la fuerza y sabiduría para vencer cualquier obstáculo que se presentó en este recorrido y no decaer en los momentos difíciles.

Gracias a la Universidad Central de Venezuela y a la facultad de ingeniería, por ser mi segunda casa, permitir mi formación profesional y brindarme gratas experiencias en estos años.

A mi familia y amigos, por estar allí y apoyarme siempre, en los buenos y malos momentos. Gracias a ellos pude superar aquellas etapas difíciles que se presentaron a lo largo de este viaje.

Quiero agradecer también a nuestra tutora, la Prof. Eliana Alvarado y al Sr. Raul Paez, por creer en nosotros y darnos la oportunidad de desarrollar este proyecto, que integra un tema que me apasiona y el mundo petrolero. Quiero darle las gracias también, por habernos dedicado parte de su tiempo para orientarnos y ayudarnos a lo largo de este arduo trabajo.

A la empresa Petrocedeño, por permitir el desarrollo de este trabajo, al brindarnos toda la información necesaria para poder llevar a cabo este trabajo de investigación.

A Osmer Parabavire, y a los profesores Carlos González y Diego Manfre, por encaminarnos y prestarnos su ayuda en los momentos en los que nos sentíamos perdidos y sin un rumbo fijo, dándonos el impulso y la orientación necesarios para continuar con este proyecto.

A mi compañera de trabajo, Adriana Mijares, a Héctor Navarro y a Lineke Morales, por convertirse en mi segunda familia y brindarme su apoyo en estos últimos años.

A los profesores de la escuela de petróleo, por transmitirnos sus conocimientos de manera dedicada y brindarnos su apoyo y amistad a lo largo de este recorrido.

A todos, mi eterna gratitud.

John Sánchez

Mijares M. Adriana A., Sánchez L. John D.

DESARROLLO DE UNA HERRAMIENTA DE ANÁLISIS DE DATOS DE PRODUCCIÓN Y OPERATIVIDAD DE UN GRUPO DE POZOS, USANDO MÉTODOS DE *DATA SCIENCE* Y *MACHINE LEARNING*

**Tutor académico: Ing. Eliana Alvarado
TEG. Caracas, Universidad Central de Venezuela. Facultad de Ingeniería,
Escuela de Petróleo, Año 2021, páginas:288**

Palabras claves: Ciencia de datos (*Data Science*) - Petróleo, aprendizaje automatizado (*Machine learning*) - Petróleo, Análisis de datos, predicción de producción, productividad de pozos, bombas de cavidades progresivas (*BCP*), Faja Petrolífera del Orinoco.

Durante las operaciones regulares en la industria petrolera, se genera una gran cantidad de información de distinta índole, la cual, es necesario procesar para poder transformarla en datos de valor que faciliten la toma de decisiones. Este TEG buscó analizar bases de datos de producción y de servicios a pozos (manejo de bombas de cavidad progresiva) en un campo que trabaja con un sistema automatizado de producción y procesos. En el desarrollo se usaron técnicas y metodologías de *Data Science*, implementando la metodología CRISP-DM, que consta de 6 etapas: comprensión del negocio, comprensión y preparación de los datos, modelado, evaluación y, por último, el despliegue. Para lograr los objetivos se diseñaron distintos algoritmos, con la finalidad de unificar bases de datos, mejorar la calidad de los datos y generar modelos a través de técnicas de *Machine Learning* para la predicción de producción y el análisis del comportamiento de bombas de cavidades progresivas. Entre las técnicas evaluadas están: la regresión lineal multivariable, regresión polinomial, redes neuronales artificiales, *support vector machine*, *K-nearest neighbors* y árbol de decisiones. La herramienta permitió procesar y analizar los datos de producción y operación de 40 pozos y determinar la técnica de *Machine Learning* con mejor ajuste. Los resultados mostraron que el modelo de predicción con mejor ajuste es el polinomial de grado 2, seguido por la red neuronal, mientras que, para el análisis del comportamiento de las bombas, la técnica con mayor precisión fue el árbol de decisiones, logrando identificar 6 tipos de problemas operacionales. El desarrollo de esta herramienta permitió predecir producción y realizar un análisis rápido y eficiente de los datos, con la finalidad de optimizar el funcionamiento de los pozos.

ÍNDICE

RESUMEN.....	ix
ÍNDICE DE TABLAS	xix
ÍNDICE DE FIGURAS.....	xxi
INTRODUCCIÓN.....	1
CAPÍTULO I.....	3
1 PLANTEAMIENTO DEL PROBLEMA	3
1.1.- PLANTEAMIENTO DEL PROBLEMA	3
1.2.- OBJETIVOS	4
1.2.1.- Objetivo general.....	4
1.2.2.- Objetivos específicos	4
1.3.- JUSTIFICACIÓN	4
1.4.- ALCANCE.....	5
1.5.- LIMITACIONES	5
CAPÍTULO II.....	7
2 MARCO TEÓRICO.....	7
2.1.- YACIMIENTOS	7
2.1.1.- Yacimientos de acuerdo con su composición.....	7
2.1.2.- Yacimientos de acuerdo con su saturación de gas.....	10
2.1.3.- Yacimientos según su compresibilidad.....	11
2.1.4.- Mecanismos de empuje de un yacimiento	11
2.1.4.1.- Empuje por agua.....	12
2.1.4.2.- Gas en solución.....	13
2.1.4.3.- Expansión de la capa de gas	13
2.1.4.4.- Segregación gravitacional.....	14

2.1.4.5.- Empuje por expansión roca-fluido.....	15
2.2.- PRODUCCIÓN DE YACIMIENTOS.....	16
2.2.1.- Principios de producción de un pozo.....	17
2.2.1.1.- Presión estática del yacimiento.....	17
2.2.1.2.- Temperatura de yacimiento	17
2.2.1.3.- Presión de fondo fluyente	18
2.2.1.4.- Presión de cabezal.....	18
2.2.1.5.- Diferencial de presión (<i>Drawdown</i>)	18
2.2.1.6.- Viscosidad.....	18
2.2.1.7.- Tasa de líquido.....	19
2.2.1.8.- Porcentaje de agua y sedimentos	20
2.2.1.9.- Relación agua petróleo	20
2.2.1.10.- Relación gas petróleo.....	21
2.2.1.11.- Pruebas de Presión (<i>Well testing</i>)	21
2.2.2.- Arreglo tipo <i>cluster</i> o macolla	23
2.2.2.1.- Elementos de una macolla en subsuelo.....	23
• Pozos	23
○ Potencial de un pozo.....	24
○ Índice de productividad	25
2.2.2.2.- Elementos e instalaciones de una macolla en superficie	25
• Múltiple de diluyente.....	26
• Válvulas multipuerto	26
• Líneas de flujo	26
• Medidor multifásico	26

• Bomba multifásica.....	27
2.2.3.- Problemas comunes de producción en el sistema pozo yacimiento	27
2.2.3.1.- Daño de formación	27
2.2.3.2.- Intrusión de agua.....	28
2.2.3.3.- Intrusión de gas.....	29
2.2.3.4.- Producción de arena.....	30
2.2.4.- Estimulación de un pozo.....	30
2.3.- SISTEMAS DE PRODUCCIÓN	31
2.3.1.- Completación de un pozo	32
2.3.2.- Bombeo de cavidades progresivas (BCP).....	32
2.3.2.1.- Conceptos básicos.....	33
• Torque	33
• Sumergencia	33
• Nivel estático y nivel dinámico	33
2.3.2.2.- Principio de funcionamiento de una BCP.....	34
2.3.2.3.- Descripción del equipo BCP.....	36
• Equipos de subsuelo	37
• Equipos de superficie:	41
2.3.2.4.- Aplicaciones.....	43
2.3.2.5.- Características operativas de las BCP.....	44
• Desplazamiento volumétrico o caudal:	44
• Altura de descarga o head.....	44
2.3.2.6.- Clasificación de las BCP.....	45

• Bombas industriales:	45
• Bombas para la industria petrolera:	45
• Clasificación de las BCP de acuerdo con su ensamblaje	46
○ Bombas tubulares	46
○ Bombas tipo insertables:	46
• Clasificación de las BCP de acuerdo con su geometría	46
○ Bombas de geometría simple o lobular	47
○ Bombas multilobulares	47
• Configuraciones especiales de BCP	47
○ <i>Charge Pumps</i>	47
○ BCP con sistema eléctrico de manejo en fondo (<i>Electric Downhole Drive PCP Systems</i>)	48
2.3.2.7.- Diagnóstico y optimización	48
2.3.2.8.- Mantenimiento de los equipos	49
2.4.- EL ROL DE LA PRODUCCIÓN DE PETRÓLEO EN LA DEMANDA ENERGÉTICA MUNDIAL.....	52
2.5.- ANTECEDENTES DEL USO DE <i>DATA SCIENCE</i>	54
2.6.- ESTADÍSTICA.....	57
2.6.1.- Normalización.....	57
2.6.2.- Estandarización.....	58
2.6.3.- Significancia estadística (<i>p-values</i>)	58
2.7.- <i>DATA SCIENCE</i>	59
2.7.1.- Metodologías de <i>data science</i>	59
2.7.1.1.- Metodología SEMMA	60

2.7.1.2.- Metodología CRISP-DM	61
2.7.2.- Disciplinas asociadas a <i>Data Science</i>	63
2.7.2.1.- Minería de datos (<i>data mining</i>).....	64
2.7.2.2.- <i>Big data</i>	64
2.7.2.3.- Inteligencia artificial (<i>artificial intelligence</i>).....	64
2.7.2.4.- Aprendizaje automatizado (<i>machine learning</i>).....	65
2.7.2.5.- Aprendizaje profundo (<i>deep learning</i>)	65
2.8.- APRENDIZAJE AUTOMATIZADO (<i>MACHINE LEARNING</i>)	65
2.8.1.- Aprendizaje supervisado.....	66
2.8.2.- Aprendizaje no supervisado.....	67
2.8.3.- Aprendizaje reforzado (<i>Reinforcement learning</i>)	67
2.8.4.- Metodologías de <i>machine learning</i>	68
2.8.4.1.- Regresión lineal multivariable	68
2.8.4.2.- Regresión polinómica	72
2.8.4.3.- <i>K-nearest neighbors</i>	73
2.8.4.4.- Árbol de decisiones.....	74
2.8.4.5.- Redes neuronales artificiales (<i>deep learning</i>).....	75
• Arquitectura básica de una red neuronal	75
○ Capas:.....	76
• Funciones internas de una RNA	77
2.9.- DESCRIPCIÓN DEL ÁREA DE ESTUDIO	82
CAPITULO III.....	89
3 MARCO METODOLÓGICO.....	89

3.1.- REVISIÓN BIBLIOGRÁFICA Y RECOPIACIÓN DE INFORMACIÓN, Y DATOS DE INTERÉS.....	89
3.2.- APLICACIÓN DE LA METODOLOGÍA CRISP-DM	91
3.2.1.- Definición de las necesidades o entendimiento del negocio (<i>Business Understanding</i>)	92
3.2.2.- Estudio y comprensión de los datos (<i>Data Understanding</i>)	93
3.2.2.1.- Estudio de las bases de datos y material suministrado para la selección de la muestra.....	93
3.2.2.2.- Estudio, análisis y organización de las bases de datos	97
3.2.2.3.- Análisis de la producción y operatividad del campo basado en gráficos.....	107
• Cálculo de variables adicionales necesarias para el análisis	108
• Unión de bases de datos y análisis del historial de producción y comportamiento de bombas	109
3.2.3.- Análisis y preparación de los datos.....	113
3.2.3.1.- Análisis, identificación y visualización gráfica de datos perdidos	114
• Identificar el formato de los datos perdidos	114
• Contabilizar datos nulos en el historial de producción.....	115
• Identificar las fechas donde se encuentran los datos nulos	116
• Estudio de los rangos de datos de cada uno de los atributos	116
• Identificar y graficar todas las anomalías y servicios de interés en el historial de producción	118
3.2.3.2.- Metodologías para la caracterización estadística.....	119
3.2.3.3.- Transformación de datos de texto a <i>Dummy data</i>	125

3.2.3.4.- Creación de <i>datasets</i> finales para el modelado	126
• Set de datos reales	126
• Set de datos normalizados	127
• Set de datos estandarizados	127
3.2.4.- Modelado y evaluación	130
3.2.4.1.- Modelado de la producción.....	130
• Modelo de regresión lineal multivariable.....	132
• Regresión polinómica.....	137
• Redes neuronales	138
3.2.4.2.- Modelado del comportamiento de las BCP	148
3.2.5.- Despliegue	154
CAPÍTULO IV	155
4 ANÁLISIS DE RESULTADOS	155
4.1.- ANÁLISIS DE LA SELECCIÓN DE LAS METODOLOGÍAS DE <i>MACHINE LEARNING</i> Y <i>DATA SCIENCE</i>	155
4.2.- ANÁLISIS DE LOS MODELOS DE PREDICCIÓN	157
4.2.1.- Regresión lineal multivariable.....	157
4.2.2.- Regresión polinomial.....	161
4.2.2.1.- Modelo de regresión polinomial de grado 2	161
4.2.2.2.- Modelo de regresión polinomial de grado 3	162
4.2.2.3.- Modelo de regresión polinomial de grado 4	163
4.2.2.4.- Análisis y comparación de los tres modelos polinomiales	164
4.2.3.- Redes neuronales	168
4.2.4.- Importancia de la caracterización estadística.....	174

4.2.5.-	Importancia de la normalización.....	176
4.2.6.-	Selección del modelo final de predicción.....	177
4.3.-	SIGNIFICANCIA ESTADÍSTICA	179
4.4.-	ANÁLISIS DEL MODELADO DEL COMPORTAMIENTO DE LAS BCP	186
4.4.1.-	Aumentos excesivos de gas	190
4.4.2.-	Cambios muy bruscos en el diferencial de presión.....	192
4.4.3.-	Aumentos excesivos de la tasa de agua	194
4.4.4.-	Cambios bruscos en el torque	196
4.4.5.-	Exceso de inyección de diluyente.....	201
4.4.6.-	Fallas críticas por pérdida de fluidos	203
4.4.7.-	Análisis final de las bombas en la macolla en estudio.....	205
4.5.-	VALIDACIÓN DE LA PRECISIÓN DEL MODELO REALIZADO CON EL MODELO ACTUAL DE SEGUIMIENTO	208
4.6.-	ETAPA DE DESPLIEGUE	211
CAPÍTULO V		212
5	CONCLUSIONES	212
6	RECOMENDACIONES.....	215
ANEXOS		216
ANEXO 1: ACRÓNIMOS		216
ANEXO 2: PROCEDIMIENTO MANUAL PARA LA OPTIMIZACIÓN DE UN POZO CON SISTEMA DE BOMBEO BCP DEL ING. NELVY CHACIN		218
ANEXO 3: PROBLEMAS TÍPICOS EN BCP Y POSIBLES SOLUCIONES HIRSCHFELDT (2008),		219
ANEXO 4: TABLA DE ESPECIFICACIONES DE BCP		227

ANEXO 5: RESULTADOS DE LA REGRESIÓN LINEAL MULTIVARIABLE (SET DE DATOS REALES CARACTERIZADOS).....	254
ANEXO 6: RESULTADOS DE LA REGRESIÓN POLINOMIAL (SET DE DATOS REALES CARACTERIZADOS).....	258
ANEXO 7: RESULTADOS DE REDES NEURONALES (SET DE DATOS NORMALIZADOS CARACTERIZADOS, FUNCIÓN DE ACTIVACIÓN MAE)	269
ANEXO 8: COMPARACIÓN DE RESULTADOS DE LOS MODELOS EN CADA POZO.....	271
ANEXO 9: RESULTADO DEL ANÁLISIS DE BOMBAS Y SU SIGNIFICANCIA ESTADÍSTICA POR POZO Y MODELO DE BOMBA.....	273
ANEXO 10: MANUAL PARA EL USUARIO DE LA HERRAMIENTA <i>AJ</i> <i>ANALYTIC TOOL</i>.....	281
BIBLIOGRAFÍA.....	304

ÍNDICE DE TABLAS

Tabla 1: Clasificación del petróleo de acuerdo con su gravedad API.....	9
Tabla 2: Comparación entre las etapas de la metodología SEMMA y CRISP-DM ...	63
Tabla 3: Análisis de datos para la selección de la muestra	97
Tabla 4: Tabla generalizada con información principal de cada uno de los pozos.....	98
Tabla 5: Descripción de los parámetros en historial de producción	99
Tabla 6: Descripción de los parámetros en la base de datos de los servicios	105
Tabla 7: Tabla de atributos finales para el modelado (Caso base).....	131
Tabla 8: Casos según el set de datos a utilizar	133
Tabla 9: Parámetros calculados para estudiar eficiencia del modelo.....	135
Tabla 10: Esquema de combinaciones RNA.....	144
Tabla 11: Configuración de parámetros para la etapa 1	145
Tabla 12: Ejemplo de configuración de parámetros para la etapa 2	146
Tabla 13: Resultados de la selección del set de datos para regresión lineal multivariable	158
Tabla 14: Errores absolutos y tiempos de predicción por arena en el método de regresión lineal multivariable.....	159
Tabla 15: Comparación de resultados en los distintos sets de datos para polinomio de grado 2.....	161
Tabla 16: Comparación de resultados en los distintos sets de datos para polinomio de grado 3.....	162
Tabla 17: Comparación de resultados en los distintos sets de datos para polinomio de grado 4.....	163
Tabla 18: Comparación de resultados de error relativo y coeficiente de correlación en los métodos de regresión polinomial.....	164
Tabla 19: Resultados de la etapa 1 en el ambiente sedimentario deltaico	169
Tabla 20: Resultados de la etapa 1 en el ambiente sedimentario fluvial	170
Tabla 21: Resultados del modelo de redes neuronales por arena.....	173
Tabla 22: Selección del mejor modelo por pozo.....	177

Tabla 23: Comparación del error relativo en los tres modelos principales..... 178

Tabla 24: Comparación del error absoluto en los tres modelos principales 178

Tabla 25: Comparación del tiempo de entrenamiento en los tres modelos principales
..... 179

Tabla 26: Comparación de significancia estadística para las arenas del ambiente
deltaico 185

Tabla 27: Comparación de significancia estadística para las arenas del ambiente fluvial
..... 186

ÍNDICE DE FIGURAS

Figura 1: Clasificación de los yacimientos según su composición (Modificado).....	8
Figura 2: (a) Ejemplo de yacimiento subsaturado (b) Ejemplo de yacimiento saturado	10
Figura 3: Ejemplo del empuje por agua en un yacimiento de petróleo.....	12
Figura 4: Ejemplo del empuje por expansión de la capa de gas en un yacimiento de petróleo.....	14
Figura 5: Ejemplo del empuje por expansión de la capa de gas en un yacimiento de petróleo.....	15
Figura 6: Ejemplo del empuje por compactación de la roca en un yacimiento de petróleo (a) Antes de la expansión (b) después de la expansión.....	16
Figura 7: Área de drenaje en un pozo horizontal.....	24
Figura 8: Intrusión o conificación de agua en un pozo.....	29
Figura 9: Conificación de gas en un pozo.....	29
Figura 10: Nivel estático y nivel dinámico de los fluidos en el pozo (Modificado)...	34
Figura 11: (a) Elementos principales de una BCP (b) Estator (c) rotor (d) Configuración de una bomba BCP Fuente: Protex – Progressive Cavity Pump Manual. pag 5.....	36
Figura 12: Principales componentes de una BCP.....	38
Figura 13: Ejemplo de un ancla de torque (Modificado).....	39
Figura 14 Gráfico de trabajos realizados en los equipos de superficie.....	51
Figura 15: Gráfica del consumo energético mundial desde el año 1993 hasta el 2018 (Modificado).	52
Figura 16: Total de reservas de petróleo del mundo.....	53
Figura 17: Comparación entre la producción real y los valores obtenidos por la red neuronal.....	56
Figura 18: Disciplinas y metodologías asociadas a Data Science (Modificado).....	63
Figura 19: Diferencia entre la programación clásica y machine learning.....	66
Figura 20: Algoritmo de regresión lineal (Modificado).....	69
Figura 21: Ejemplo de Regresión lineal Simple.....	70

Figura 22: Regresión lineal multivariable (Modificado)	71
Figura 23: Comparación entre las ecuaciones de regresión (Modificado).....	72
Figura 24: Ejemplo de K-Nearest Neighbors.....	74
Figura 25: (a) Arquitectura de una neurona biológica (B) Arquitectura de una neurona artificial	75
Figura 26: Ejemplo de buen ajuste, comportamiento ideal.....	80
Figura 27 (a) Ejemplo de una curva de entrenamiento subentrenada (b) ejemplo de una curva con sobreajuste	81
Figura 28: Ubicación del área de estudio en la FPO Bloque Junín.....	83
Figura 29: (a) Corte transversal del campo en estudio (b) Ejemplo de canalización del acuífero en los pozos horizontales del campo.....	84
Figura 30: Distribución de las macollas y troncales en el campo en estudio.....	85
Figura 31: (a) Arreglo de una macolla de pozos horizontales en paralelo (b) arreglo de una macolla tipo estrella.....	86
Figura 32: Diagrama de una Macolla.....	86
Figura 33: Completación típica de un pozo en el campo en estudio.....	87
Figura 34: Etapas de la metodología CRISP-DM	91
Figura 35: Mapa de la distribución de las macollas en el campo Petrocedoño.....	94
Figura 36: Ubicación del sensor de fondo encargado de medir la presión de salida de la bomba y la presión de revestidor	102
Figura 37: Tabla de Clasificación de Servicios por pozo de la macolla en estudio..	107
Figura 38: Gráfica interactiva para análisis de datos por parámetro en el pozo VD20	110
Figura 39: Gráfica interactiva para comparar y estudiar el comportamiento de más de un parámetro en un periodo de tiempo determinado del pozo VD20.	111
Figura 40: Análisis del comportamiento de las bombas presentes en el pozo VD20	113
Figura 41: Esquema de datos perdidos	114
Figura 42: Ejemplo de resultados al contabilizar los datos nulos en el pozo VD20.	115
Figura 43: Ejemplo de datos nulos presentes en el historial de producción del pozo VD20.....	116

Figura 44: Estudio del periodo inicial del pozo VD20	117
Figura 45: Ejemplo de estudio de anomalías en el historial de producción en el pozo VD20.....	118
Figura 46: Ejemplo de caracterización estadística por valor medio en el pozo VD20 (a) Parámetro sin caracterizar (b) Parámetro caracterizado.....	120
Figura 47: Ejemplo de caracterización estadística por valor promedio en el pozo VD20 (a) Parámetro sin caracterizar (b) Parámetro caracterizado	121
Figura 48: Ejemplo de caracterización estadística por regresión lineal en el pozo VD20 (a) Parámetro de presión de entrada sin caracterizar (b) Parámetro de presión caracterizado	122
Figura 49: Ejemplo de caracterización estadística por el método de restauración de presión en el pozo VD20 (a) Parámetro de presión de entrada sin caracterizar (b) Parámetro de presión caracterizado	124
Figura 50: Ejemplo de separación de datasets de entrenamiento y prueba.....	129
Figura 51: Esquema base del modelo de regresión lineal multivariable.....	134
Figura 52: Ejemplo del modelo de predicción de regresión lineal multivariable del pozo VD20.....	136
Figura 53: Ejemplo del resultado de la predicción del pozo VD20 por regresión polinómica comparando los grados del polinomio.	138
Figura 54: Estructura base de la RNA con tres capas ocultas.....	142
Figura 55: Selección de las mejores RNA en el pozo VD20.....	146
Figura 56: Esquema de la metodología para definir la configuración óptima de parámetros de una RNA.	147
Figura 57: Tabla de descripción de las bombas en el pozo VD20.....	149
Figura 58: Calculo de tasa ideal en la bomba 1 del pozo VD20.....	151
Figura 59: (a) Gráfica resultante del estudio del comportamiento bomba 1 del pozo tipo	152
Figura 60: Resultado de la predicción de producción pozo VD20, con el método de regresión lineal multivariable.....	160
Figura 61: Estudio del modelo de regresión polinomial en la arena B2	165

Figura 62: Resulta del modelo de regresión polinomial del pozo VD20. (a) Regresión polinomial grado 2 (b) regresión polinomial grado grado?	167
Figura 63: Resultado del modelo de regresión polinomial grado 4pozo VD20.....	168
Figura 64: Resultados de la etapa 1 (Metodología redes neuronales) (a) Selección del set datos optimo (b) Selección de función de pérdida.....	171
Figura 65: Resultados de la etapa 2 (Metodología redes neuronales) (a) Selección del escalamiento	172
Figura 66: Resultado del modelo de redes neuronales en el pozo VD20.....	174
Figura 67: Análisis de resultados de los tres modelos de acuerdo con la calidad del dato	175
Figura 68: Análisis de resultados de los tres modelos respecto a la normalización .	176
Figura 69: Significancia estadística de los atributos en la arena B2	181
Figura 70: Significancia estadística de los atributos en la arena C2	182
Figura 71: Significancia estadística de los atributos en la arena D3	183
Figura 72: Significancia estadística de los atributos en la arena E1	184
<i>Figura 73: (a) Análisis del comportamiento de la bomba 1 en el pozo tipo respecto a la predicción.....</i>	<i>188</i>
<i>Figura 75: (a) Respuesta de la tasa de petróleo por un aumento de gas en el pozo</i>	<i>191</i>
<i>Figura 76: (a) Respuesta de la tasa de petróleo por cambio del diferencial de presión en el pozo VD03</i>	<i>193</i>
<i>Figura 77: (a) Respuesta de la tasa de petróleo por producción de agua en el pozo VD02</i>	<i>195</i>
<i>Figura 78: (a) Respuesta de la tasa de petróleo por cambios bruscos en el torque posible producción de sólidos en el pozo VD06</i>	<i>197</i>
<i>Figura 79: (a)Comportamiento de la producción de petróleo bomba 2 pozo VD35.</i>	<i>200</i>
<i>Figura 80: (a)Comportamiento de la producción de petróleo bomba 3 pozo VD20.</i>	<i>202</i>
<i>Figura 81: (a)Comportamiento de la producción de petróleo bomba 1 pozo VD18.</i>	<i>204</i>
<i>Figura 82: Análisis de cambios de bombas innecesarios</i>	<i>205</i>
<i>Figura 83: Análisis de cambios de bombas necesarios</i>	<i>206</i>
<i>Figura 84: Análisis de cambios de bombas de acuerdo con el modelo.....</i>	<i>207</i>

Figura 85: Análisis del comportamiento de los modelos de bombas respecto a su duración y productividad.....	208
Figura 86: (a) Comportamiento del modelo de predicción en el año 2002 -2013 (b)Comportamiento del modelo de predicción en el año 2014 -2018.....	210
Figura 87: Panel inicial de la interfaz gráfica	211

INTRODUCCIÓN

En los últimos años se ha observado un aumento notable en el uso de tecnologías como *Machine Learning* y *Data Science* en diversas áreas. Importantes empresas petroleras como Shell, TransOcean, SBM Offshore, Equinor y Devon Energy han aumentado su productividad considerablemente, al analizar grandes cantidades de datos en los diferentes departamentos y detectar tendencias que les facilitan la toma de decisiones. Los algoritmos de *Machine Learning* permiten analizar un grupo de datos de entrada con sus respectivas respuestas, para establecer una función o modelo que se ajuste al grupo de datos y luego aplicar ese modelo para estudiar las respuestas de otros datos de entrada. (Harris C. 2014)

En la industria petrolera, se maneja gran cantidad y variedad de datos provenientes de los procesos de exploración y explotación de hidrocarburos, tales como: registros de pozos, análisis de núcleos y fluidos, datos de producción, servicios, entre otros. Frecuentemente, los datos son almacenados para ser analizados posteriormente por el departamento encargado, con el fin, de predecir o estimar, algunos parámetros que no se pueden cuantificar directamente (como las reservas o la declinación de un yacimiento) y que normalmente requiere mucho esfuerzo, experiencia y tiempo por parte de los ingenieros, para ser llevado a cabo de una manera satisfactoria (Primera A. y otros, 2016; Johnston J y Guichard A. 2015).

Gracias al desarrollo de *Data Science* y *Machine Learning*, en conjunto con sistemas computarizados más rápidos y poderosos, ahora es posible analizar una gran cantidad y variedad de datos (bien sean datos de producción, operatividad, datos sísmicos, registros eléctricos, datos de perforaciones, entre otros) en un tiempo relativamente corto en comparación con la manera en que se realizaban normalmente, es decir, que el equipo encargado invirtiera horas o días incluso recolectando y preparando los datos para su posterior análisis.

Durante el desarrollo de este TEG se aplica la metodología de análisis de datos CRISP-DM, abarcando 6 fases: el entendimiento del negocio, estudio y comprensión de los datos, el análisis y la preparación de los datos, el modelado, la evaluación y por último el despliegue. Para la fase de modelado se realiza un modelo probabilístico de predicción, basado en tres técnicas de *machine learning* (regresión lineal multivariable, regresión polinómica y redes neuronales), los cuales permiten predecir la producción de petróleo en muy pocos segundos y con bajos errores. Sin embargo, es necesario estudiar dos características principales (calidad del dato y normalización) para determinar el modelo de predicción óptimo en cada pozo. Adicionalmente se utiliza otra técnica de *machine learning* (árbol de decisiones) que permite determinar los parámetros que afectan en la productividad del pozo y en el comportamiento del sistema de levantamiento (bombas de cavidades progresivas).

El resultado de la investigación es una herramienta de optimización de análisis de datos de producción y operación para pozos petroleros, las cuales, incluye todas las fases de la metodología CRISP-DM, y que permite realizar un análisis robusto de diferentes parámetros, que intervienen durante la producción y operación de los pozos.

La herramienta permite detectar en poco tiempo, fallas y anomalías que pueden ser determinantes para la toma de decisiones, incluyendo: realizar cambios operativos, mantenimiento preventivo a los equipos, servicios a los pozos o estimulaciones según sea el caso.

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1.- Planteamiento del problema

Durante las operaciones de exploración y producción en la industria petrolera, es necesario el análisis de una gran cantidad de datos que se obtienen mediante los diversos registros y mediciones que se realizan durante las operaciones. Registros sísmicos, petrofísicos, datos de laboratorio, mediciones de presión y producción, son solo algunos de los tipos de datos que se deben analizar, con el fin de tomar decisiones que optimicen las operaciones relacionadas con la industria de hidrocarburos.

La gran cantidad y diversidad de información a analizar hacen que se le dedique una gran parte del tiempo a la recolección, preparación y estudio de datos, muchas veces omitiendo parte de ellos para agilizar el proceso. El tiempo que se emplea para la preparación de los datos pudiera utilizarse para analizar y tomar decisiones o resolver algún problema que se presente durante las distintas operaciones que se llevan a cabo en la industria, bien sea en la exploración, perforación o producción. Adicionalmente, al omitir datos para facilitar y agilizar el proceso de análisis, es posible que se pueda pasar por alto información que ayude a describir algún patrón o tendencia, que describa mejor el comportamiento de algunos parámetros de interés.

Actualmente, se han desarrollado herramientas y métodos computacionales como *Machine Learning*, *Data Science* y *Big Data*, que reducen de manera significativa el tiempo del flujo de trabajo y que adicionalmente, pueden procesar y analizar una gran cantidad y variedad de datos. En consecuencia, no es necesario omitir información que puede resultar valiosa, por el contrario, se pueden procesar grandes cantidades de información si la plataforma tecnológica lo permite.

1.2.- Objetivos

1.2.1.- Objetivo general

Desarrollar una herramienta de optimización de análisis de datos de producción y operación de un grupo de pozos usando *Data Science* y *Machine Learning*.

1.2.2.- Objetivos específicos

- Revisar la información bibliográfica acerca de herramientas de análisis de datos como *Data Science* y *Machine Learning* y su aplicación en la industria petrolera.
- Analizar los datos disponibles de producción y operación de pozos petroleros para seleccionar técnicas y metodologías de *Data Science*.
- Desarrollar un software para la preparación y el procesamiento de los datos usando *Machine Learning*, para generar un modelo que describa la tendencia a lo largo del tiempo de los parámetros de un grupo de pozos en estudio.
- Validar la precisión del modelo generado con el modelo actual de seguimiento de los pozos en estudio a través de parámetros de producción y operación.
- Comparar la eficiencia de la metodología planteada con datos históricos para la predicción de parámetros de producción.

1.3.- Justificación

En la industria petrolera, algunos de los recursos más importantes son la administración del tiempo y el resguardo de los datos como principal activo. En la administración del dato, se dedica una gran cantidad de tiempo a la recolección y preparación de todos los datos para su posterior análisis. Dicha tarea se complica cuando la cantidad de datos a analizar es muy grande, y por lo tanto se requiere una mayor cantidad de horas-hombre de trabajo para el equipo encargado de esa tarea, lo

que se traduce en mayores costos para la compañía a cargo. Al desarrollar una herramienta que pueda preparar y procesar todos los datos necesarios para el análisis de manera automática y más eficiente, la cantidad de tiempo que normalmente se requiere para completar esa labor se reduciría considerablemente, reduciendo también los costos asociados y optimizando el análisis e interpretación de los mismos, para la toma de decisiones oportuna y eficiente.

En los pozos de estudio seleccionados, el actual método de análisis y seguimiento genera grandes retrasos y altos tiempos de espera desde la carga de los datos provenientes del campo hasta los formatos de análisis previamente definidos por la operadora. Por lo tanto, no hay una adecuada visualización, ni manejo de éstos en las diferentes áreas de trabajo, generando retrasos en las actividades derivadas de estos análisis, toma de decisiones no fundamentadas en datos reales, así como los planes de servicios a pozos, entre otros. Es por eso que las herramientas destinadas a este proceso pueden ser mejoradas y optimizadas mediante el uso de *Data Science* y *Machine Learning* a fin de reducir tiempos y costos.

1.4.- Alcance

En el desarrollo de este trabajo, el análisis de los datos disponibles permitirá proponer una metodología en el área de *Data Science* y junto a técnicas de *Machine Learning*, desarrollar un software que prepare y procese los datos de operatividad y producción en un corto período de tiempo. Luego, usando técnicas de visualización, proveer una manera más sencilla de interpretar y entender los resultados obtenidos y poder hacer recomendaciones al respecto.

1.5.- Limitaciones

La principal limitación de este trabajo es no poder contar con una mayor cantidad de datos de operatividad y producción de los pozos en estudio, pues para poder implementar las técnicas de *Machine Learning* de una manera óptima, se debe ajustar

el modelo con una gran cantidad de información para que los resultados sean más satisfactorios. Adicionalmente, solo se van a procesar pocos parámetros de operación y producción, debido a la manera que los datos fueron registrados, siendo lo ideal, trabajar con la mayor cantidad de datos que se puedan, pues de esa manera, se podrían identificar tendencias o patrones que normalmente no se podrían reconocer. Finalmente, si se desea procesar una gran cantidad y variedad de datos, es necesaria una plataforma de desarrollo distinta, dado que se estaría trabajando con métodos de *Big Data*, tópico que no se trabaja en esta investigación.

CAPÍTULO II

MARCO TEÓRICO

El marco teórico de esta investigación está dividido en 8 secciones principales, las cuales comprende, conceptos básicos de los yacimientos, la producción de yacimientos y los sistemas de producción, antecedentes de la ciencia de datos aplicados a la industria petrolera, conceptos básicos de estadística, ciencia de datos y aprendizaje supervisado y por último una breve descripción del área en estudio.

Todas las secciones aquí explicadas son fundamentales para el desarrollo de esta investigación.

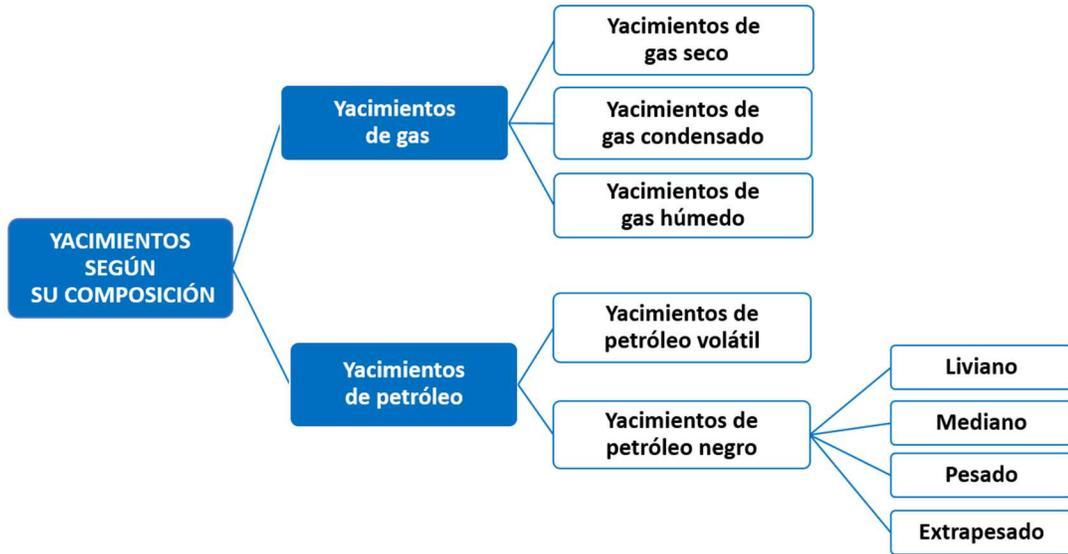
2.1.- Yacimientos

Un yacimiento de hidrocarburos se define como el lugar donde se encuentra de forma natural una acumulación de petróleo o gas ^[1]. Estos yacimientos presentan tres propiedades básicas de clasificación necesarias para la planificación de un plan de explotación y producción. Los yacimientos se pueden clasificar de acuerdo con su composición, de acuerdo con su saturación de gas y según su compresibilidad.

2.1.1.- Yacimientos de acuerdo con su composición

Un yacimiento se puede clasificar de acuerdo con el tipo de hidrocarburo que se encuentre presente en la roca. Este criterio de clasificación se basa en el estado de la materia, composición química, diagrama de fases, temperaturas, presiones, densidad, viscosidad, entre otros. En este caso se clasifican en dos grandes bloques: yacimientos de gas y yacimientos de petróleo. En la figura 1 se muestra un esquema completo de la clasificación de los hidrocarburos de acuerdo con su composición. Los yacimientos de gas se dividen en tres: yacimientos de gas seco, gas húmedo y gas condensado, mientras los yacimientos de petróleo se dividen en dos grupos principales: petróleo volátil y

petróleo negro. Este último presenta una subdivisión en cuatro subgrupos: petróleo liviano, mediano, pesado y extrapesado, la cual se basa en la densidad o gravedad API que presente el petróleo [2].



*Figura 1: Clasificación de los yacimientos según su composición (Modificado)
Fuente: Clase de ingeniería de yacimientos (Padrón A.)*

Este TEG, está enfocado en un yacimiento de petróleo negro pesado y extrapesado, el cual se describe más adelante.

Los yacimientos de petróleo negro se pueden definir de acuerdo con sus características y su diagrama de fases. Estos se caracterizan por: presentar temperaturas por debajo de la temperatura crítica, poseer mayor cantidad de compuestos pesados (C_{7+}), su peso molecular va desde 70 lb/mol y puede superar las 210 lb/mol, presentan relación gas petróleo menor a los 2000 PCN/BN, su factor volumétrico de formación es de hasta 1,5 BY/BN y su gravedad API es menor a los 40° [2]. Como se menciona con anterioridad, los yacimientos de petróleo negro se clasifican a su vez en cuatro subgrupos de acuerdo con su gravedad API, los cuales se muestran en la tabla 1, siendo la clasificación de petróleo pesado y extrapesado los de mayor interés en este TEG.

Tabla 1: Clasificación del petróleo de acuerdo con su gravedad API

CLASIFICACIÓN DEL PETRÓLEO	GRAVEDAD API
Livianos	30° - 40°
Medianos	22° - 30°
Pesados	10° - 22°
Extrapesados	0° - 10°

Una de las propiedades principales del petróleo es su capacidad de fluir, la cual se mide a través de la viscosidad y su unidad de referencia es el centipoise (cp). La viscosidad es representada indirectamente por la densidad o gravedad específica en la escala de °API. El petróleo pesado se encuentra en el rango de 10° a 22° API, mientras el petróleo extrapesado se encuentra en el rango de los 0° a 10° API, ambos subgrupos comprenden una viscosidad entre 1.200 y 95.000 cp y dependen de los elementos con los que estén compuestos.

La viscosidad es una propiedad fundamental para la producción, tratamiento y manejo del yacimiento, por lo tanto, para hacer que los fluidos fluyan de manera más fácil, se requieren métodos de recuperación terciaria como calentamiento o inyección de diluentes. Otra característica importante de este tipo de crudos es que tienen una alta concentración de azufre, sal y metales, todo esto ocasiona que los equipos de producción y refinación estén en constante mantenimiento. Adicionalmente es necesario aplicar métodos y tratamientos especiales para mejorar su calidad con la finalidad de obtener los resultados deseados de comercialización^[3]. La composición del petróleo es una propiedad fundamental, sin embargo, este tipo de clasificación no es la única que debe tomarse en cuenta al momento de realizar el plan de explotación del yacimiento.

2.1.2.- Yacimientos de acuerdo con su saturación de gas

Los yacimientos de petróleo se pueden clasificar en dos tipos de acuerdo con su saturación inicial de gas: yacimientos saturados y yacimientos subsaturados.

Los yacimientos saturados son aquellos donde la presión inicial del yacimiento es menor o igual a la presión de burbuja. Este tipo de yacimientos se considera un yacimiento bifásico que contiene una fase líquida y otra gaseosa. Por otro lado, se puede decir que un yacimiento está saturado con gas, cuando el yacimiento se encuentra a presión y temperatura constante y al momento de reducir ligeramente la presión, éste, libera gas en solución.

Un yacimiento es subsaturado cuando la presión inicial del mismo es mayor que la presión de burbuja. En este tipo de yacimientos el gas se encuentra disuelto en el petróleo, por lo tanto, no existe capa de gas. Se determina que un yacimiento es subsaturado cuando éste se encuentra a presión y temperatura constante, y al aplicarse una reducción de la presión, éste no libera gas y la fase líquida permanece constante.

La figura 2a representa un yacimiento subsaturado en el cual, sólo existe la fase líquida. Una vez este tipo de yacimientos alcanza la presión de burbuja, el gas es liberado formando una capa de gas y el yacimiento pasa a ser un yacimiento saturado como se muestra en la figura 2b^[4].

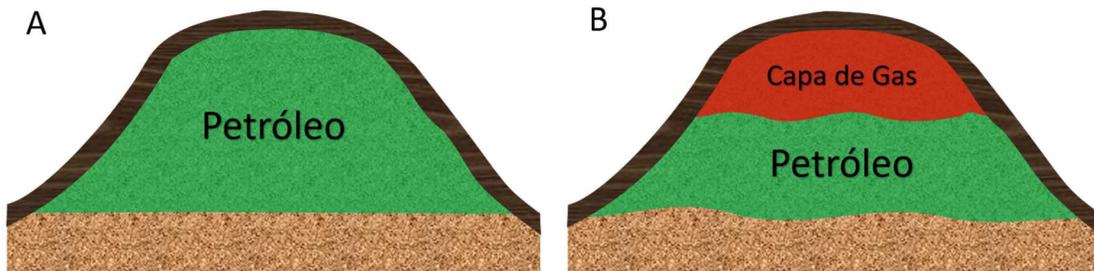


Figura 2: (a) Ejemplo de yacimiento subsaturado (b) Ejemplo de yacimiento saturado

2.1.3.- Yacimientos según su compresibilidad

Por último, se tiene la clasificación de los yacimientos según su compresibilidad, donde existen dos tipos de yacimientos: consolidados y no consolidados.

Los yacimientos consolidados están compuestos por sedimentos, los cuales han atravesado diferentes procesos geológicos como la cementación, el metamorfismo, entre otros, generando una roca sedimentaria sólida. A través de este tipo de sedimentos, los fluidos fluyen a través de fracturas o del espacio poroso sin arrastrar casi ningún sedimento o fino. Se caracterizan por ser rocas como conglomerados, areniscas, limolitas, pizarras, piedras calizas y el carbón. ^{[5][6][7]}

Los yacimientos no consolidados están compuestos por sedimentos o materiales sueltos, no compactados o cementados, que pueden ser arcillas, arenas y gravas. Los fluidos en este tipo de yacimientos fluyen a través de los espacios vacíos entre los granos y arrastran con ellos parte de los sedimentos. ^{[5][6][7]}

Estos tres parámetros de clasificación deben ser tomados en cuenta para el desarrollo del plan de explotación y producción del yacimiento, sin embargo, existen otros conceptos básicos necesarios para entender el comportamiento de los mismos, como lo son los mecanismos de empuje.

2.1.4.- Mecanismos de empuje de un yacimiento

Los mecanismos de empuje son los encargados de aportar la energía requerida al sistema con el fin de que los fluidos puedan desplazarse dentro del yacimiento. Es muy importante conocer estos mecanismos previos a la producción del campo, con la finalidad de realizar un diseño óptimo de las instalaciones de superficie y la completación de los pozos. Existen cinco mecanismos de empuje natural: empuje por

agua, capa de gas, gas en solución, compactación de la roca y segregación gravitacional^[8].

2.1.4.1.- Empuje por agua

Este tipo de empuje es considerado el mecanismo natural más eficiente, logrando una recuperación de hasta un 60%. El empuje de agua se caracteriza por estar asociado a un acuífero, aportando altas presiones al yacimiento. Por lo general, el agua del acuífero se expande y desplaza al petróleo o al gas desde el yacimiento al pozo como se muestra en la figura 3. Este tipo de mecanismo requiere importante atención en el contacto agua petróleo, al igual que entre la producción de petróleo y el volumen de agua en movimiento dentro del yacimiento, ya que el exceso de agua puede generar problemas operacionales en el pozo. ^{[8][9]}

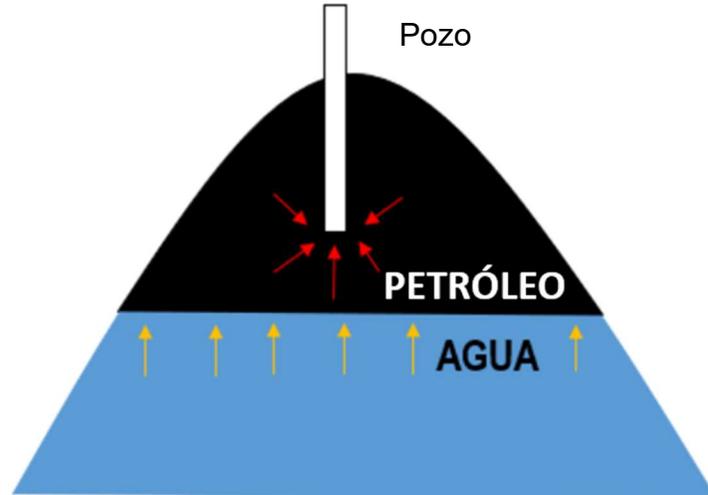


Figura 3: Ejemplo del empuje por agua en un yacimiento de petróleo
Fuente: Mecanismos de empuje. Pamela Montserrat Cortés Torres.
<http://epmex.org/news/2020/08/13/mecanismos-de-empuje/>

2.1.4.2.- Gas en solución

El gas en solución actúa como un mecanismo de empuje en yacimientos subsaturados, donde la temperatura y la presión se encuentran por encima del punto de burbuja. El factor de recobro de este tipo de mecanismo va desde un 15% a un 25%. En este tipo de yacimientos, a medida que se presenta el agotamiento de la presión, el gas se expande dentro del crudo sin liberarse ninguna burbuja al sistema. Ésta expansión del gas en solución permite un empuje del petróleo hacia el pozo. ^{[8][9]}

2.1.4.3.- Expansión de la capa de gas

La expansión de la capa de gas se presenta en yacimientos saturados, donde la temperatura y la presión se encuentran por debajo del punto de burbuja. En este tipo de yacimientos existe un equilibrio entre el gas libre y el petróleo presente. A medida que aumenta la producción de petróleo, la presión del sistema disminuye, mientras parte del gas en solución se libera y parte del gas presente en la capa superior se expande, ambos actúan como un pistón impulsando el petróleo al pozo como se muestra en la figura 4. ^{[8][9][10][11]}

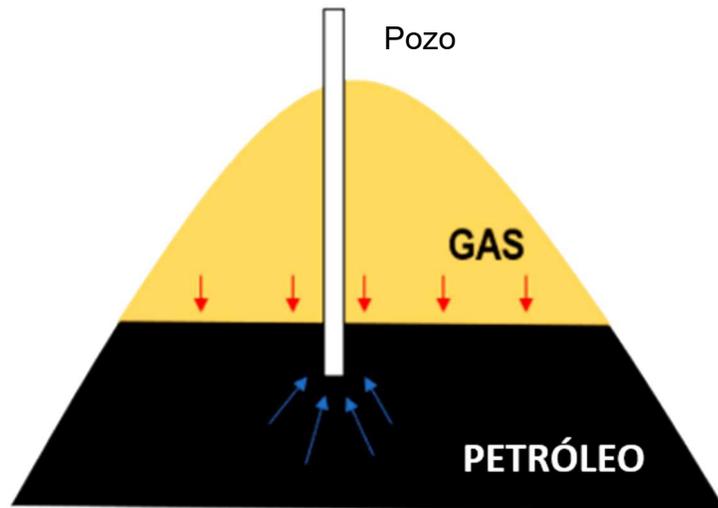


Figura 4: Ejemplo del empuje por expansión de la capa de gas en un yacimiento de petróleo

*Fuente: Mecanismos de empuje. Pamela Montserrat Cortés Torres.
<http://epmex.org/news/2020/08/13/mecanismos-de-empuje/>*

2.1.4.4.- Segregación gravitacional

Este tipo de mecanismos presenta un factor de recobro de aproximadamente un 40 a 80%. Se caracteriza por ser un mecanismo de empuje lento y se presenta cuando las fuerzas gravitacionales son mayores a las fuerzas viscosas dentro del yacimiento. En yacimientos con capa de gas que presenten una buena permeabilidad vertical, el gas tiende a moverse hacia arriba por la diferencia de densidades, actuando como un mecanismo de empuje y generando que el petróleo se mueva hacia a la parte baja del yacimiento. ^{[9][11]}

Por otro lado, también se presenta el drenaje gravitacional, el cual ocurre en yacimientos con gran buzamiento, donde el comportamiento tanto del gas como del petróleo es similar al antes mencionado, sin embargo, el flujo es paralelo al ángulo de buzamiento como se muestra en la figura 5. ^{[8][11]}

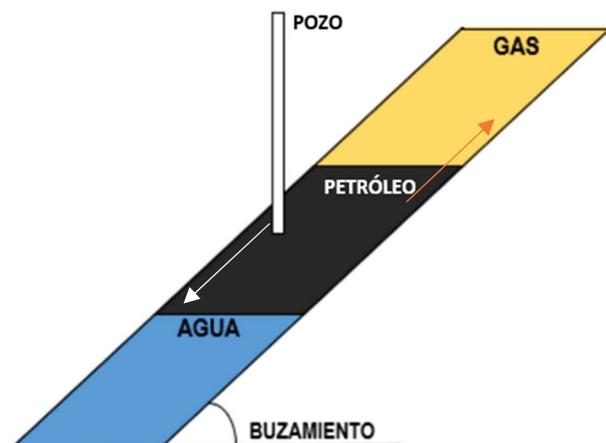
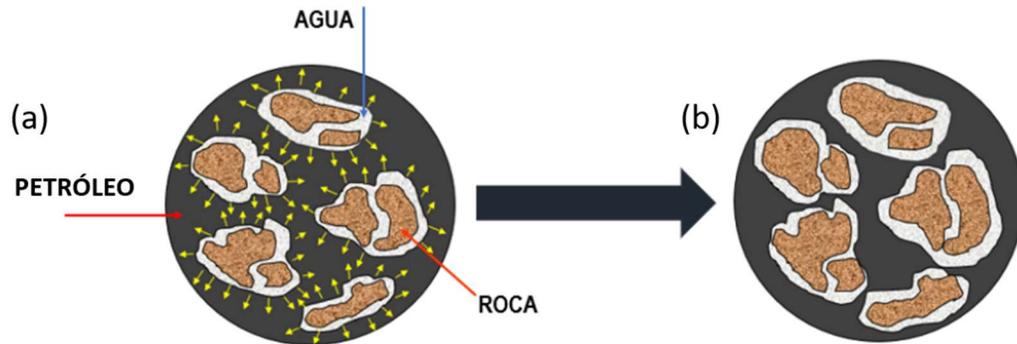


Figura 5: Ejemplo del empuje por expansión de la capa de gas en un yacimiento de petróleo

*Fuente: Mecanismos de empuje. Pamela Montserrat Cortés Torres.
<http://epmex.org/news/2020/08/13/mecanismos-de-empuje/>*

2.1.4.5.- Empuje por expansión roca-fluido

El empuje por expansión roca-fluido es un mecanismo de empuje común en todos los yacimientos. Las rocas suprayacentes se mantienen ejerciendo una presión constante en el sistema, generando la compactación entre los granos y la expansión de los fluidos. A medida que la presión del yacimiento disminuye y los fluidos son producidos, los sedimentos de la roca tienden a reorganizarse y los fluidos remanentes se expanden dentro del sistema para ocupar los espacios vacíos.^[11] En la figura 6 se muestra un ejemplo de este tipo de mecanismos de empuje. La figura 6a representa en color amarillo la separación existente entre los granos antes del proceso de expansión, mientras que la figura 6b muestra como el espacio entre los granos disminuye y los fluidos existentes alrededor de los granos se expanden.



*Figura 6: Ejemplo del empuje por compactación de la roca en un yacimiento de petróleo
 (a) Antes de la expansión (b) después de la expansión
 Fuente: Mecanismos de empuje. Pamela Montserrat Cortés Torres.
<http://epmex.org/news/2020/08/13/mecanismos-de-empuje/>*

Los mecanismos de empuje son fundamentales para la producción de cualquier yacimiento y deben formar parte de las estrategias de explotación.

Hasta este punto se ha definido el sistema yacimiento, su clasificación y los mecanismos de empuje, pero, para poder engranar toda esta investigación es necesario conocer los principios de producción del yacimiento.

2.2.- Producción de yacimientos

La producción de yacimientos es un proceso complejo que depende de muchas variables, como lo son: los principios de producción, los elementos y equipos tanto de subsuelo como de superficie necesarios para la producción de los hidrocarburos y los problemas más comunes que se presentan al iniciar la producción de hidrocarburos en un yacimiento.

2.2.1.- Principios de producción de un pozo

La producción de hidrocarburos en un pozo se estudia con base a distintos parámetros del yacimiento, como la presión, la temperatura del yacimiento y los parámetros medidos en el pozo (la presión de fondo y de cabezal). De igual forma se deben tomar en cuenta las propiedades de los fluidos como la viscosidad, la densidad, entre otros. Existen otros principios importantes como las características operacionales (tasas de producción, porcentaje de agua y sedimentos (%AyS), relación agua petróleo y relación gas petróleo), y las pruebas *well testing*, como *build up*, *drawdown*, la prueba de inyectividad y la prueba *falloff*.

Todos estos parámetros se definen a continuación:

2.2.1.1.- Presión estática del yacimiento

La presión estática (P_e) se define como la presión que tiene un yacimiento a condiciones iniciales cuando no existe movimiento de fluidos dentro del mismo y todas las fases se encuentran en equilibrio, sin embargo, la presión estática puede variar a medida que se extraen los fluidos, donde se considera como una presión promedio del yacimiento. ^{[4][12]}

2.2.1.2.- Temperatura de yacimiento

La temperatura del yacimiento es la magnitud que mide la cantidad de calor con la que se encuentran los fluidos o hidrocarburos a nivel de yacimiento. A medida que los hidrocarburos son producidos, éstos llegan a la superficie con una temperatura menor, esta temperatura se conoce como temperatura de cabezal. ^[13]

2.2.1.3.- Presión de fondo fluyente

La presión de fondo fluyente (P_{wf}) es la presión medida en el fondo del pozo una vez éste se encuentre en producción y está gobernada por la presión ejercida desde superficie. ^[4]

2.2.1.4.- Presión de cabezal

La presión de cabezal o también conocida *Wellhead pressure* (P_{wh}) se define como la presión que puede ser controlada desde superficie, con el fin de controlar las presiones del fondo. ^{[4][13]}

2.2.1.5.- Diferencial de presión (*Drawdown*)

Se define como la diferencia que existe entre la presión estática y la presión del fondo fluyente y se puede representar con la siguiente ecuación: ^[13]

$$\Delta P = P_e - P_{wf} \quad \text{Ec. 1}$$

Donde:

P_e = Presión estática

P_{wf} = Presión fondo fluyente

2.2.1.6.- Viscosidad

La viscosidad es una característica de todos los fluidos, tanto líquidos como gases y se define como la resistencia o la oposición interna de un fluido al flujo debido a los efectos combinados de cohesión y adherencia. Es un parámetro fundamental ya que afecta de forma directa los procesos de producción, transporte, refinación y petroquímica, ^[14]

La temperatura, la presión y la densidad son parámetros que ejercen importantes efectos sobre la viscosidad:

- Efecto de la temperatura sobre la viscosidad

La temperatura es una propiedad que genera un efecto notable en la viscosidad, su efecto es diferente en muestras de petróleo líquido que en gases. En el caso de los gases, al aumentar la temperatura la viscosidad de los gases aumenta, mientras que, en el petróleo, si la temperatura aumenta la viscosidad disminuye. Un aumento de temperatura en el crudo origina una disminución en la fuerza de cohesión y en la resistencia de las moléculas a desplazarse. ^[14]

- Efecto de la presión sobre la viscosidad

La presión mecánica genera un aumento en la viscosidad, siempre y cuando no exista una adición de gas en el hidrocarburo. Un aumento de presión genera un cambio en la compresión del sistema, esto ocasiona una disminución de la distancia entre las moléculas aumentando la resistencia de las moléculas a desplazarse. ^[14]

- Efecto de la densidad sobre la viscosidad

La densidad está relacionada con la flotabilidad de un fluido, por lo tanto, entre más denso sea un fluido mayor será su viscosidad. ^[14]

2.2.1.7.- Tasa de líquido

La tasa de líquido (Ql) es la sumatoria entre la tasa de petróleo producido y la tasa de agua. Es considerado un parámetro fundamental para el estudio de las bombas en esta investigación y se calcula con la siguiente ecuación: ^[13]

$$Q_l = Q_o + Q_w \quad Ec. 2$$

Donde:

Ql= Tasa de líquido

Qo= Tasa de petróleo

Qw= Tasa de agua

2.2.1.8.- Porcentaje de agua y sedimentos

El porcentaje de agua y sedimentos (%AyS) se define como la cantidad de agua y sedimentos producida con respecto a la tasa total de líquido y se expresa con la siguiente ecuación:^[13]

$$\%AyS = \frac{Q_w}{Q_l} * 100 \quad Ec. 3$$

Donde:

Ql= Tasa de líquido

Qw= Tasa de agua

%AyS= Porcentaje de agua y sedimentos

2.2.1.9.- Relación agua petróleo

La relación agua petróleo se define como la división de la tasa de agua con respecto a la tasa petróleo. Es considerado uno de los parámetros más importantes al momento de estudiar el potencial del pozo y los problemas comunes en el sistema de producción. Esta relación se calcula con la siguiente ecuación:^[14]

$$RAP = \frac{Q_w}{Q_o} \quad Ec. 4$$

Donde:

RAP= Relación agua petróleo

Qw= Tasa de agua

Qo= Tasa de petróleo

2.2.1.10.- Relación gas petróleo

Es la relación entre la tasa de gas producido con respecto a la cantidad de petróleo producido. Se considera otro parámetro importante en la investigación, ya que depende del gas en solución y tiene una relación directa con los mecanismos de empuje presentes en el área. Su ecuación es la siguiente:^{[13][14]}

$$RGP = \frac{Q_g}{Q_o} \quad Ec. 5$$

Donde:

RGP= Relación gas petróleo

Qg= Tasa de gas

Qo= Tasa de petróleo

2.2.1.11.- Pruebas de Presión (*Well testing*)

Las pruebas de presión son pruebas que se le realizan a un pozo durante un periodo de tiempo, en el cual se induce una perturbación al sistema (se altera la tasa de producción) con la finalidad de estudiar el comportamiento de algunos parámetros como presión y flujo, bien sea, en el fondo o en el cabezal del pozo. Si la prueba se realizó correctamente, el análisis de estos parámetros permite estimar algunas propiedades del yacimiento, como el factor de daño (*skin factor*), el almacenamiento, la presión estática, el área de drenaje, entre otros. ^{[15][16]}

Existen varios tipos de pruebas de presión como:

- *Drawdown*: en esta prueba, un pozo que permaneció cerrado por un tiempo prolongado es abierto y se deja que fluya a una tasa constante. [8]
- Prueba de inyektividad o de inyección: esta prueba es muy similar a la prueba *drawdown*, con la diferencia que en vez de dejar que los fluidos fluyan fuera del pozo, se inyectan fluidos a una tasa constante. [8]
- Prueba *Falloff*: En esta prueba, un pozo que estaba sometido a un flujo de inyección con tasa constante se cierra, y se estudia el comportamiento de la presión. [8]
- Prueba de interferencia: en estas pruebas, un pozo es producido mientras que en otro pozo cercano se miden las alteraciones de la presión, debido a la producción del primero. [8]
- *BuildUp*: en esta prueba, un pozo que este fluyendo (idealmente a una tasa constante) se cierra, y se mide la presión de fondo. [8]

Todos los parámetros antes mencionados son fundamentales para entender el comportamiento del pozo a través de los datos, sin embargo, existen otros elementos que son fundamentales, como lo son los arreglos de pozo y los equipos de producción. Una estrategia de explotación debe tomar en cuenta el tipo de yacimiento y plantear distintos escenarios de perforación y ubicación de los pozos para encontrar la forma óptima de drenar los hidrocarburos. Por último, es necesario crear un sistema de producción que se ajuste a las características de los fluidos presentes y que faciliten la extracción de éstos desde el pozo hasta la superficie.

Los pozos pueden ser diseñados y perforados de forma vertical u horizontal. En la actualidad se han implementado diferentes arreglos de pozo en los campos, estos pueden ser: en líneas paralelas a las estructuras, en arreglos tipos para recuperación secundaria y en arreglos tipo *cluster* o macollas. Estos últimos son de mucho interés para esta investigación, ya que, es el arreglo presente en el campo de estudio que definiremos más adelante.

2.2.2.- Arreglo tipo *cluster* o macolla

Una macolla o *cluster* se define como un grupo de pozos unidos que operan de forma individual, sin embargo, a nivel de superficie pueden ubicarse en una misma zona, donde cada uno de los pozos están conectados a válvulas multipuertos.

Entre los arreglos de pozos más comunes se presentan pozos paralelos, los cuales se perforan separados a una distancia establecida de acuerdo con un estudio previo, con el fin de que no exista interferencia en sus áreas de drenaje hasta este límite.

2.2.2.1.- Elementos de una macolla en subsuelo

A nivel de subsuelo una macolla está constituida por pozos, los cuales son perforados definiendo un arreglo específico.

- Pozos

Un pozo se define como el medio que permite la producción de los fluidos desde el yacimiento hasta la superficie. Desde los inicios de la exploración petrolera todos los pozos se construían con trayectorias verticales, sin embargo, con el pasar de los años, los avances tecnológicos han permitido el desarrollo de la perforación direccional controlada.^[11]

En yacimientos de petróleo pesado y extrapesado es recomendable perforar pozos horizontales, ya que permiten tasas de flujo más altas con una caída de presión menor y además minimizan los problemas de producción de arena. Los pozos horizontales son perforados con un ángulo de inclinación no menor a 86° respecto a la vertical o paralelos a los estratos del yacimiento con el fin de tener una mayor área de producción.^[11] En este tipo de pozos las condiciones de flujo de los fluidos son diferentes a un pozo vertical, ya que en este tipo de pozos se crea un área de drenaje en forma elipsoidal como se muestra en la figura 7, la cual permite mayor cantidad de

petróleo producido por pozo al final de su vida productiva, reduciendo el número de pozos requeridos para desarrollar un yacimiento.

Estos pozos por lo general son equipados con una bomba electrosumergible o una bomba de cavidades progresivas que permiten levantar el petróleo hasta la superficie. La productividad de este tipo de pozos dependerá del tipo de fluido, de la longitud y de la completación del mismo. Todos estos parámetros son fundamentales

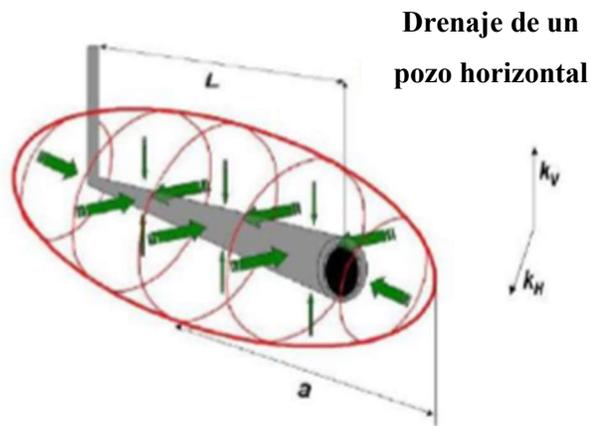


Figura 7: Área de drenaje en un pozo horizontal

Fuente: Miranda Lisbeth, Producción en pozos horizontales (2016)

en el desarrollo del yacimiento y su factibilidad económica, por esta razón, es necesario conocer cuál es el potencial de un pozo y de cuales parámetros depende:

- Potencial de un pozo

El potencial de un pozo se define como la tasa de flujo que es capaz de producir teóricamente un pozo, el cual puede ser medido a través del índice de productividad que se define a continuación.

- Índice de productividad

El índice de productividad es una medida del potencial o la capacidad que tiene un pozo de producir fluidos y se define como la relación entre la tasa producida de un fluido con respecto a un diferencial de presión o *drawdown* dado y se puede calcular con la siguiente ecuación: ^[17]

$$IP = \frac{Q}{P_e - P_{wf}} \quad Ec. 6$$

Donde:

Q= Tasa de fluido en estudio

Pe= Presión estática

Pwf= Presión de fondo fluyente

IP= Índice de productividad

Todos los elementos antes mencionados, como la presión de fondo fluyente asociada al diferencial de presión y otros elementos que afecten la producción deben controlarse desde la superficie, por lo que, en una macolla son necesarios distintos elementos e instalaciones de superficie.

2.2.2.2.- Elementos e instalaciones de una macolla en superficie

Las instalaciones de superficie dependen de cada campo. En la zona de estudio, una macolla para producción está compuesta a nivel de superficie, de instalaciones de monitoreo y elementos que permiten el correcto funcionamiento y la posibilidad de bombear los fluidos a la estación central. Entre los elementos importantes de una macolla están: los múltiples de diluyente, las válvulas multipuertos, líneas de flujo, el medidor multifásico y la bomba multifásica; mientras que en las instalaciones de superficie se encuentran: el cuarto de control de la macolla, el cuarto de control del sistema de levantamiento y el cuarto de control de la bomba multifásica. ^[10]

- Múltiple de diluyente

El múltiple de diluyente permite realizar la inyección de diluyente en los pozos seleccionados con la finalidad reducir la viscosidad del crudo y aumentar los niveles de recobro. ^[10]

- Válvulas multipuerto

Es un sistema compacto de recolección de producción, el cual cuenta con dos salidas, una salida al sistema de producción continuo y una salida de aislamiento de la producción direccionado al medidor multifásico. Con esta configuración es posible conectar los pozos y aislar uno de ellos alineándolo con la línea de prueba, todos los demás se mantienen conectados y produciendo a través del cabezal de producción y fluyendo por la salida común hacia la bomba multifásica. ^[18]

- Líneas de flujo

Las líneas de flujo son las tuberías que permiten mantener la conexión del sistema y transportar los fluidos desde el pozo hasta la bomba multifásica o al medidor. ^[10]

- Medidor multifásico

Un medidor multifásico es un dispositivo que permite registrar las tasas de flujo de fluidos individuales de petróleo y gas cuando pasa más de un fluido a través de una línea de conducción. Éste brinda lecturas exactas, incluso cuando existen regímenes de flujo diferentes. ^[19]

- Bomba multifásica

El objetivo principal del bombeo multifásico consiste en bombear fluidos inestables, como: mezclas de petróleo, gas, agua y sedimentos, provenientes de los pozos sin separación previa, éstos fluidos son bombeados por un oleoducto o troncal hacia el centro principal de separación y tratamiento de fluidos. [20]

En la siguiente sección se explica con detalle los problemas comunes que pueden ocurrir tanto a nivel de yacimiento como en el pozo los cuales pueden generar pérdidas en la producción de petróleo.

2.2.3.- Problemas comunes de producción en el sistema pozo yacimiento

En los pozos al igual que en los yacimientos existen problemas comunes como lo son el daño a la formación, la intrusión de fluidos, la producción de arena y problemas mecánicos en el pozo o en la bomba. Éstos últimos se mencionarán más adelante, cuando se explique el sistema de producción.

2.2.3.1.- Daño de formación

El daño de formación se define como una restricción al flujo de los fluidos en el medio poroso causado por la reducción de la permeabilidad en las zonas cercanas al pozo. [16]

Existen diferentes operaciones durante las cuales se puede generar un daño en la formación, como lo son: la perforación, el proceso de cementación, la completación, limpiezas y reparaciones de pozos, inyección de fluidos a la formación, la producción de un pozo, entre otros. Durante todas estas operaciones el yacimiento interactúa con elementos, fluidos y cambios de presión atípicas para su sistema. Todas estas generan una disminución en la permeabilidad, cambios en la mojabilidad o en la permeabilidad

relativa, obstrucción o taponamiento de los poros, emulsiones o cambios en la composición química de los fluidos. ^[16]

2.2.3.2.- Intrusión de agua

En la actualidad existen muchos yacimientos limitados por rocas saturadas con agua, los cuales reciben el nombre de acuíferos. En la mayoría de los casos el tamaño de los acuíferos es mucho mayor al tamaño de los yacimientos de hidrocarburos, por lo que, el acuífero se considera de extensión infinita y tiene un efecto importante sobre el yacimiento. Sin embargo, existen acuíferos pequeños en comparación con el yacimiento, donde su comportamiento resulta insignificante. ^[21]

El comportamiento de los acuíferos es importante en la producción, ya que, ante una caída de presión en el yacimiento, éste actúa para contrarrestar la declinación de la presión y genera una intrusión de agua en la roca, la cual puede deberse a: expansión de agua, expansión de otras acumulaciones de hidrocarburos, compresibilidad de la roca del acuífero y flujo artesiano, el cual se considera cuando el acuífero se eleva por encima del yacimiento. ^[21] Las intrusiones de agua son muy comunes en pozos petroleros donde existe un acuífero presente. También se conocen como conificaciones y se generan cuando el nivel de agua alcanza la zona de producción del pozo. En la figura 8, se observa un ejemplo de cómo se produce la intrusión de agua en un pozo. Este caso se puede estudiar en el historial de producción al observar un aumento significativo en la relación agua-petróleo. ^{[21][10]}

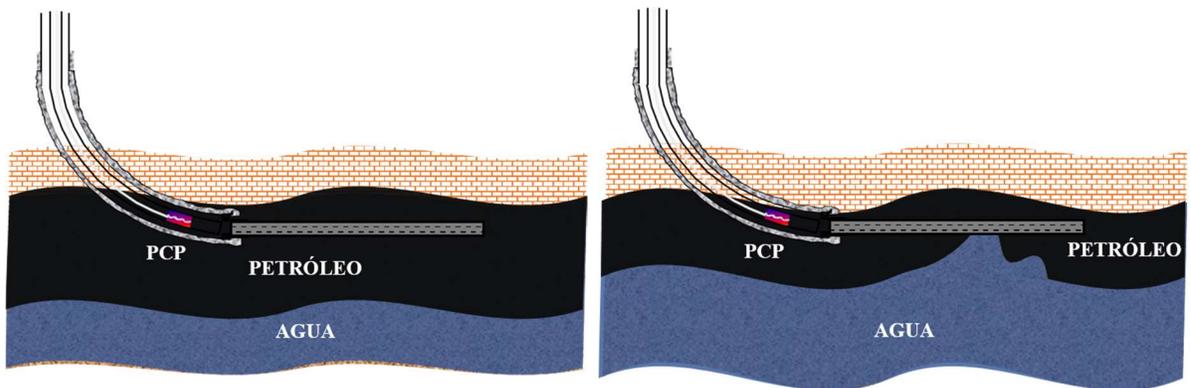


Figura 8: Intrusión o conificación de agua en un pozo
Fuente: Statoil Petrocedeño. Procesos básicos de producción. (Octubre 2010.)

2.2.3.3.- Intrusión de gas

La intrusión de gas en pozos petroleros se da en yacimientos saturados cuando la presión de burbuja está por debajo de la presión del yacimiento. En este tipo de yacimientos empieza a formarse una capa de gas que tiende a expandirse a medida que disminuye la presión, y se agotan los hidrocarburos. El gas tiende a desplazarse frente al banco de petróleo generando una mayor producción de gas, como se muestra en la figura 9. Este fenómeno podría apreciarse en un aumento significativo de la RGP, sin embargo, se debe estudiar con detalle para lograr concluir que éste es el problema presente en el pozo.

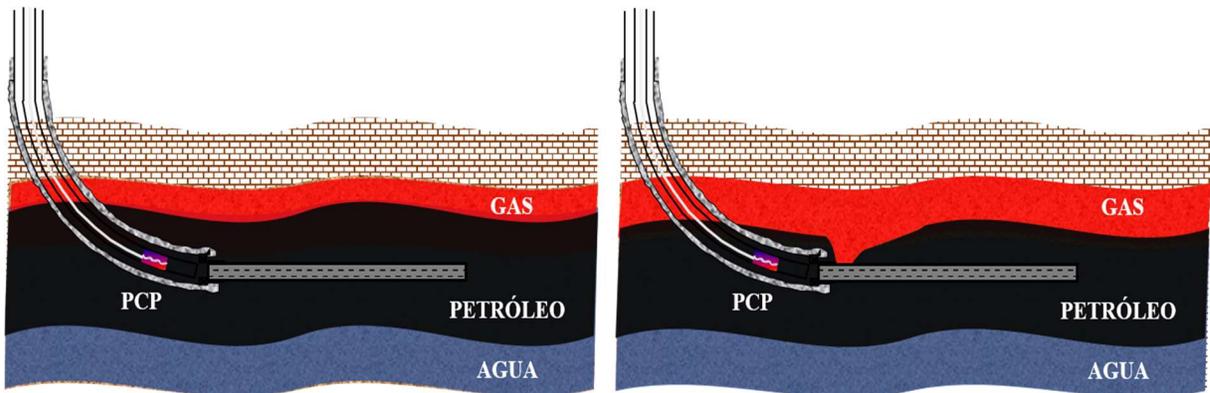


Figura 9: Conificación de gas en un pozo
Fuente: Statoil Petrocedeño. Procesos básicos de producción. (Octubre 2010.)

2.2.3.4.- Producción de arena

La producción de arena es un problema común en formaciones no consolidadas. Por lo general la producción de fluidos siempre viene acompañada de sedimentos, lo que genera una menor recuperación de los hidrocarburos, ocasiona daños en los equipos de superficie y en los equipos de fondo del pozo, así, como costos de mantenimientos elevados en el sistema. Éstos depósitos de sólidos pueden ocasionar obstrucción parcial o completa en las líneas de flujo, aumentando la corrosión y la erosión en las tuberías y en la bomba. [22]

Los problemas asociados a la intrusión de fluidos en el pozo, al igual que la producción de arena, pueden ser controlados con trabajos adicionales de completación o cambiando las condiciones operativas a nivel de superficie. Sin embargo, es necesario realizar un estudio a fondo del problema que se presente para encontrar la solución adecuada. Por otro lado, cuando existe un daño de formación se puede mejorar la permeabilidad del yacimiento a través de un proceso de estimulación, el cual se explica en la siguiente sección.

2.2.4.- Estimulación de un pozo

Se define como estimulación al tratamiento realizado a un pozo con el fin de mejorar la productividad del mismo, y consiste en la inyección de fluidos de tratamiento al yacimiento. En la actualidad, las estimulaciones se dividen en dos grandes grupos, los tratamientos de fracturamiento hidráulico y los tratamientos matriciales. [23]

- Fracturamiento hidráulico: consiste en inyectar un fluido en el yacimiento a mayor presión que la presión de fractura, creando canales de flujo altamente conductivos entre el sistema pozo yacimiento. [23]

- Tratamientos de acidificación: este tipo de tratamientos se realizan a una presión inferior a la de fractura del yacimiento, con la finalidad de restaurar la permeabilidad natural del yacimiento en las áreas cercanas al pozo. [23]

El funcionamiento del sistema pozo yacimiento es clave, pero, aún existen más elementos importantes que complementan la producción de hidrocarburos. Los sistemas de producción son aquellos que permiten transportar los fluidos desde el subsuelo hasta la superficie. En la siguiente sección se explica el funcionamiento de éstos y cuáles son los elementos que lo conforman.

2.3.- Sistemas de producción

Un yacimiento de petróleo se puede producir de dos formas, por producción por flujo natural o por producción asistida con un sistema de levantamiento artificial.

En el caso de la producción por flujo natural o también conocida como recuperación primaria, la presión del yacimiento es mucho mayor que la presión en el fondo del pozo e incluso mayor al diferencial de presión entre el fondo del pozo y las instalaciones de superficie. Como consecuencia de esto, los fluidos en el yacimiento se desplazan al interior del pozo y llegan a superficie sin la necesidad de energía adicional. Muchas veces cuando un yacimiento produce por flujo natural, con el tiempo pierde energía o presión consecuencia de la producción de petróleo a través de los pozos, por lo tanto, los fluidos no pueden vencer las presiones hasta la superficie y es necesario colocar un sistema de levantamiento artificial.

En el caso de la producción asistida o con levantamiento artificial, la diferencia de presión entre el fondo del pozo y las instalaciones de superficie no es suficiente para transportar los fluidos desde el fondo, por lo tanto, se requiere de una fuente externa de energía para poder transportar los fluidos hasta la superficie. En estos casos, es necesaria la instalación de un sistema de bombeo que pueda proporcionar la energía necesaria para llevar los fluidos presentes en el yacimiento hasta el cabezal del pozo.

Este tipo de mecanismos de producción busca la optimización técnica y económica de todo el proceso de producción del sistema yacimiento-pozo, sin perder de vista la seguridad de este. Algunos de los métodos de levantamiento artificial más comunes son, bombeo mecánico, bombeo hidráulico, bombeo electrosumergible, levantamiento por gas (*Gas-Lift*) y bombas de cavidades progresivas.

Esta investigación está basada en el funcionamiento y optimización de bombas de cavidades progresivas, sin embargo, antes de explicar el funcionamiento de éstas, es necesario definir en qué consiste la completación de un pozo.

2.3.1.- Completación de un pozo

La completación de un pozo es el proceso de instalación de todos los equipos necesarios que permitan el flujo de los fluidos desde el pozo hasta la superficie, incluyendo tuberías, sensores, revestidores, sistemas de levantamiento artificial, entre otros.

2.3.2.- Bombeo de cavidades progresivas (BCP)

A finales de los 20 e inicios de los 30, el francés René Moineau desarrolló una serie de bombas helicoidales, que después se convertirían en lo que hoy se conoce como bombas de cavidades progresivas (BCP). [24]

La BCP es uno de los métodos de levantamiento artificial más usado para producir crudos pesados y extrapesados, aunque su uso se ha extendido a pozos de petróleo liviano, así como también a pozos con altos cortes de agua. Su diseño sencillo y la ausencia de válvulas o partes reciprocantes, disminuye la probabilidad de atascamiento y el desgaste de piezas, haciendo que su instalación y mantenimiento sea simple y económico. [24]

Antes de comenzar a estudiar el funcionamiento de las bombas y sus componentes es necesario conocer conceptos básicos como el nivel estático, nivel dinámico y sumergencia.

2.3.2.1.- Conceptos básicos

- Torque

El torque es el nombre que reciben las fuerzas de torsión generadas por una BCP, el torque del sistema está compuesto por dos elementos: el torque hidráulico y el torque por fricción:

Torque hidráulico: es la fuerza necesaria para levantar el fluido y es función del desplazamiento de la bomba, la presión por fricción y la presión por el nivel dinámico.

Torque por fricción: es generado por la sarta de cabillas girando en el fluido, el cual es en función de la velocidad de rotación, el diámetro y la longitud de la sarta de cabillas. Adicionalmente existe otro torque el cual es generado por la fricción entre el rotor y el estator, el cual se obtiene en una prueba de banco.

Basado en esto el torque se puede calcular como la suma de los dos torques antes mencionados, como se observa en la ecuación número 7.

$$\text{Torque} = \text{Torque hidraulico} + \text{torque por friccion} \quad \text{Ec. 7}$$

- Sumergencia

La sumergencia (H) se define como la diferencia entre la distancia vertical de succión de la bomba (PB) y el nivel dinámico de los fluidos. ^[13]

- Nivel estático y nivel dinámico

El nivel estático (NE) es el nivel de fluido que se encuentra dentro del pozo cuando la bomba está detenida. La presión hidrostática ejercida por este fluido aunado a la presión en la tubería de revestimiento genera un equilibrio con la presión del yacimiento. ^[13]

El nivel dinámico (ND) se presenta al arrancar la bomba cuando el fluido comienza a subir por la tubería de producción hacia la superficie. Como consecuencia de esto, el nivel de fluido en el anular disminuye y con éste, la presión de fondo, ocasionando que los fluidos del yacimiento fluyan hacia el pozo. A medida que disminuye el nivel de fluido en el anular, aumenta la afluencia de los hidrocarburos al pozo. El nivel dinámico se estabiliza cuando la producción del yacimiento se iguala al caudal de la bomba. [13]

En la figura 10, se muestra la diferencia entre el nivel dinámico y el nivel estático del fluido, es importante tomar en cuenta que este nivel siempre es medido desde la superficie.

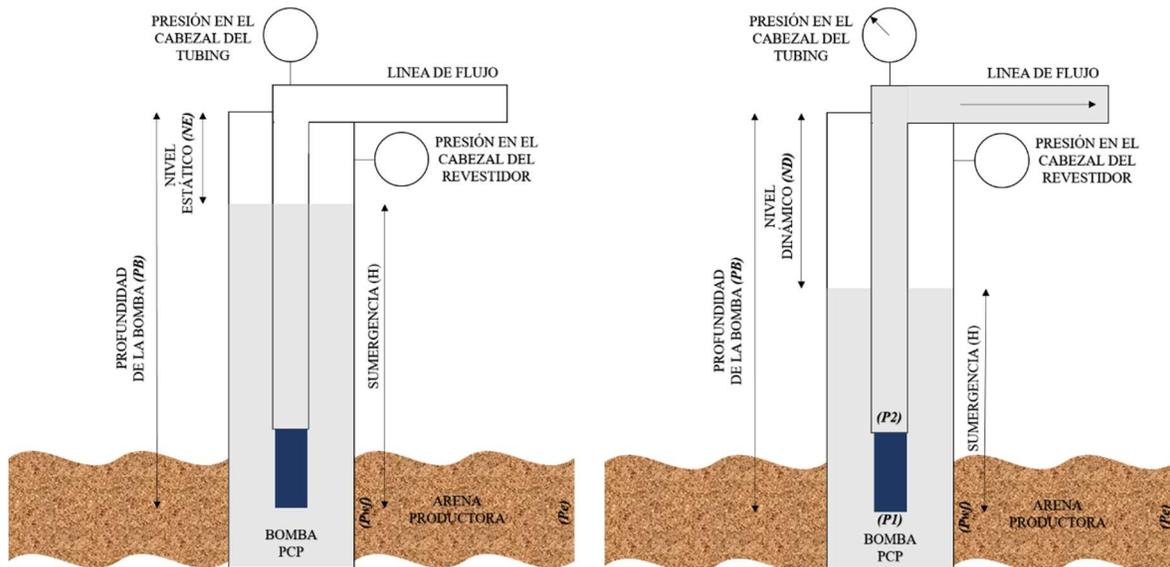


Figura 10: Nivel estático y nivel dinámico de los fluidos en el pozo (Modificado)
Fuente: Ing. Chacín Nely, Bombeo de Cavidad progresiva diciembre 2003. Documento .PDF

2.3.2.2.- Principio de funcionamiento de una BCP

En principio, una bomba de cavidades progresivas está compuesta por dos piezas principales: el rotor y el estator. El estator es el componente estático de la bomba, mientras que el rotor es el componente móvil de la misma. El rotor consiste en una pieza metálica de n lóbulos, que tiene forma helicoidal, mientras que el estator

consiste en una tubería de acero, que esta revestida en su interior por un elastómero (goma), que, al igual que el rotor, también está moldeado en forma de hélice enfrentadas entre sí. [24][13]

El funcionamiento de estas bombas se basa en el principio ideado por el francés René Moineau. El rotor se coloca dentro del estator, y debido a su geometría, se generan unas cavidades selladas en el interior del estator. Este sello se origina gracias a que las dimensiones del rotor y del estator están diseñadas para que existe una pequeña interferencia entre ellas, de manera que esta interferencia funcione como una barrera entre una cavidad y otra. En vista que las cavidades están selladas, el tipo de bombeo se considera de desplazamiento positivo. Adicionalmente, el rotor y el estator no son concéntricos, y el movimiento del rotor es combinado (uno rotacional sobre su propio eje y el otro rotacional de manera excéntrica), las cavidades se mueven de manera axial desde la entrada del motor hasta la salida. En la figura 11a, se muestra una bomba BCP con sus dos elementos principales, rotor y estator y una vista frontal de su comportamiento según el desplazamiento representado, en la figura 11b se muestra la configuración y dimensiones del estator, mientras en la figura 11c se muestra la configuración y dimensiones del rotor, y por último, en la figura 11d, se muestran las cavidades y el paso de los fluidos a través de las mismas. [24][13]

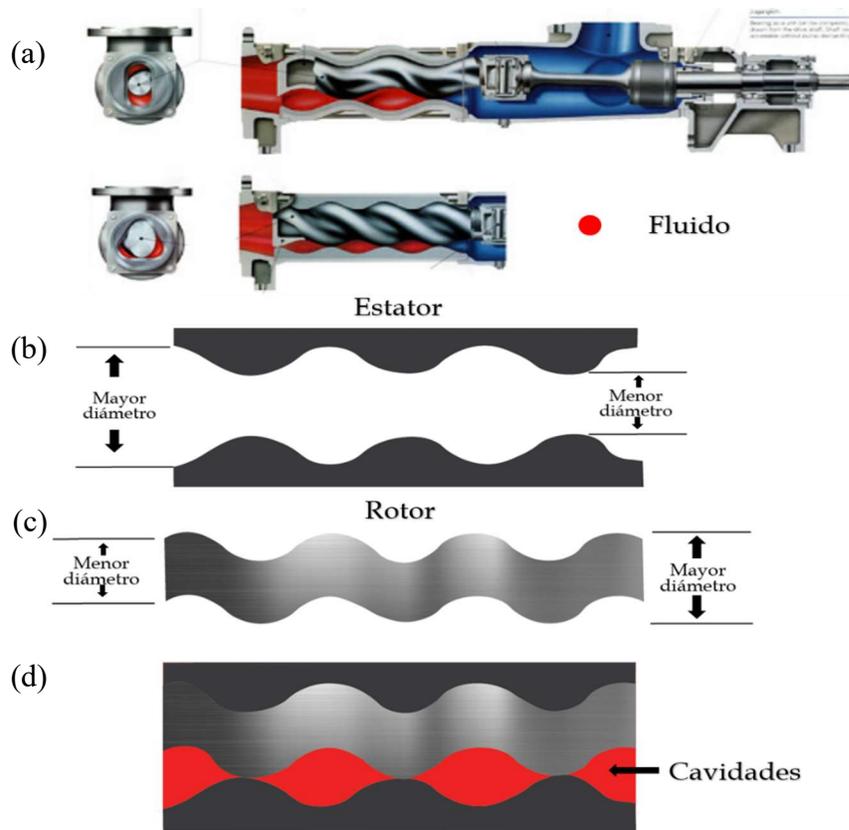


Figura 11: (a) Elementos principales de una BCP (b) Estator (c) rotor (d) Configuración de una bomba BCP
 Fuente: Protex – Progressive Cavity Pump Manual. pag 5

En la siguiente sección se describe detalladamente cada uno de los equipos y elementos que conforman una BCP.

2.3.2.3.- Descripción del equipo BCP

El equipo se puede dividir en dos secciones: los equipos de subsuelo y las instalaciones en superficie.

- Equipos de subsuelo

Son todos aquellos componentes del sistema de bombeo que operan dentro del pozo.

- **Estator:** el estator consiste en un cilindro de acero, que en su interior está revestido de un polímero de alto peso molecular (llamado elastómero), moldeado de manera helicoidal, cuyo número de lóbulos es igual al número de lóbulos del rotor más uno, esto con el fin de poder generar las cavidades. En las bombas de tipo tubular, el estator se baja al pozo junto a la tubería de producción, mientras que, en las bombas de tipo insertable, este componente se baja mediante una sarta de cabillas. ^{[13][24]}

- **Elastómero:** es un polímero del alto peso molecular, que tiene la capacidad de deformarse y recuperarse de manera elástica, dicha propiedad es la que hace posible que se genere la interferencia entre el rotor y el estator. Este componente se considera el más delicado de todo el sistema, ya que, si bien debe ser resistente químicamente para poder manejar los fluidos producidos, es susceptible a ciertos tipos de fluidos que pueden estar presentes en el yacimiento. Dichos fluidos podrían ocasionar hinchamiento, desgaste prematuro del elastómero o incluso el desprendimiento del elastómero y del cilindro metálico del estator. Los elastómeros más usados en la industria son de base nitrílica (convencionales), de hidrogenación catalítica (elastómeros hidrogenados) y los fluoelastómeros. ^{[18][25]}

- **Rotor:** es la pieza interna de la bomba de cavidades progresivas. Consiste en una pieza de acero en forma helicoidal, que posee una cubierta resistente a la abrasión, usualmente se recubre mediante un proceso electroquímico de cromado. El rotor puede ser de geometría monolobular (un solo lóbulo) o multilobular (dos o más lóbulos) según se requiera. Los procesos de fabricación de los rotores deben realizarse de manera muy cuidadosa, ya que sus dimensiones permiten un menor o mayor ajuste de la interferencia. ^{[18][25]}

En la figura 12, se observa la distribución de los componentes antes mencionados, como lo son el rotor, estator y el elastómero.

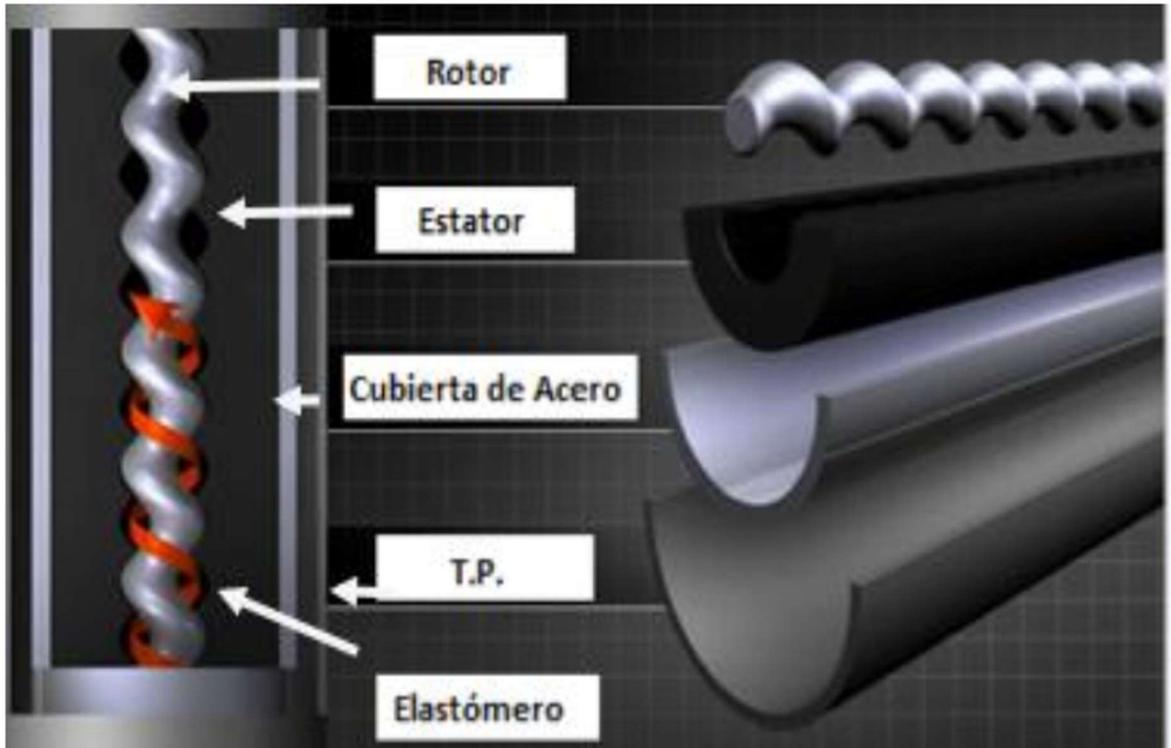


Figura 12: Principales componentes de una BCP

Fuente: Manual del participante – Sistemas artificiales de producción – PEMEX. Pag 46

- **Niple de paro:** consiste en una pequeña tubería que se coloca debajo del estator y que cumple diversas funciones, entre las que destacan:
 - ✓ En caso de rotura o desprendimiento del rotor o de la sarta de cabillas, evita que éstas caigan al fondo del pozo.
 - ✓ Permite que el rotor tenga un espacio libre durante la operación normal del sistema, debido a la posible elongación de la sarta de cabillas.
 - ✓ Servir de límite al rotor mientras se realiza el proceso de espaciado.
 - ✓ Funciona como punto de conexión para otros accesorios. ^{[13][24]}
- **Niple de maniobra:** este accesorio permite el movimiento de la cabeza del rotor, en conjunto con la conexión a la primera cabilla dentro de la tubería de

producción, debido a que el movimiento del rotor posee una excentricidad, que es mayor a su diámetro. Adicionalmente, este niple permite que, durante las operaciones de bajada o subida de la bomba, las herramientas como cuñas o mordazas se coloquen sobre el niple y no sobre el estator, lo que evitaría cualquier posible daño. [13]

- **Empacadura:** es un accesorio que se activa de manera mecánica o hidráulica, y su función es cerrar el espacio anular entre la tubería de producción y el revestidor. [13]

- **Ancla de torque:** éstas anclas evitan que la tubería se desconecte de la bomba cuando exista un torque muy fuerte al momento de la producción. Este accesorio se coloca debajo del niple de paro y se fija a la tubería de revestimiento mediante el uso de cuñas verticales. Una vez que la bomba se pone en funcionamiento, el torque que se genera hace que las cuñas se fijen al revestidor, lo que evita el giro del estator. En la figura 13, se puede observar un ejemplo de un ancla de torque, donde se observa en la parte intermedia de ella, las cuñas verticales. [13]



Figura 13: Ejemplo de un ancla de torque (Modificado)
Fuente: Manual BCP – Operación diagnóstico y fallas ESP oil pág. 33

- **Ancla de tubería:** consiste en un dispositivo que sirve para fijar la tubería de producción al revestidor, evitando así el desplazamiento o rotación de la sarta de producción. La diferencia con la empacadura es que el ancla no sella el espacio entre las dos tuberías, lo que permite la circulación de fluidos por el espacio anular. [13]

- **Anclas de gas:** son separadores estáticos que se colocan al fondo del pozo y sirven para separar el gas libre del líquido, ya que la presencia de gas libre dentro de la bomba ocasiona que la bomba pierda eficiencia volumétrica. Una de las

desventajas de estos elementos es que además de separar el gas libre del líquido, también ocasionan una caída de presión adicional, lo que induce una liberación adicional de gas y hace que el gas libre se expanda. Es necesario analizar dichos factores al momento del diseño del sistema. ^[13]

- **Centralizadores de cabillas:** estos accesorios se suelen colocar en pozos con altas desviaciones, y sirven para mantener la sarta de cabillas lo más centrada posible dentro del pozo. ^[13]

- **Niples de drenaje:** Estos dispositivos se usan cuando es necesario desalojar la tubería de producción de todo el crudo que se encuentra dentro de él (por ejemplo, cuando la sarta de cabillas falle y se necesite pescar el rotor de la bomba). La mayoría de estos nipples se pueden activar aplicando presión interna en la tubería de producción. ^[13]

- **Niples x:** estos nipples se usan para poder detectar algunas fallas (como huecos o uniones defectuosas) en la tubería de producción mediante una prueba de presión que se realiza cuando se baja la tubería de producción. Estos nipples se colocan encima del estator de la bomba. ^[13]

- **Caño filtro:** este componente se usa para evitar que, en caso de rotura o desprendimiento del elastómero, los trozos de este queden dentro del espacio anular, cuando ocurre una falla de este tipo, los trozos del elastómero pueden ser recuperados con un equipo de *pulling*, después de que la instalación de fondo haya sido reemplazada. ^{[13] [24]}

- **Trozo de maniobra:** es un accesorio que se instala por encima del rotor, en sustitución de la varilla, cuando el motor opera a más de 250 revoluciones por minuto (RPM). Este accesorio, al ser más corto que la varilla, tiene menos posibilidades de doblarse o de rozar con las paredes del rotor debido al movimiento excéntrico del mismo. ^[24]

- **Varillas de bombeo API:** son varillas de acero que están unidas entre sí mediante conexiones enroscadas, y van desde la bomba hasta la superficie, y son las encargadas de transmitir el movimiento de rotación desde la superficie hasta el rotor de la bomba. ^[24]

- **Varillas de bombeo no convencionales:** son varillas que tienen características diferentes a las varillas API y que presentan algunas ventajas, entre ellas se tienen las varillas huecas (o *hollow rods*) que, junto a una conexión tipo *Premium*, ofrecen una mayor capacidad para transmitir torque que las varillas API. Otro tipo de varillas son las varillas continuas, al no tener cuplas, se pueden usar varillas de un diámetro mayor, y, por lo tanto, tienen un menor desgaste debido al roce entre la varilla y el *tubing*. [24]

- **Vástago:** es una pieza que se coloca en la parte superior de la sarta de cabillas, el cual va empaquetado en la superficie, mediante un elemento “prensa”, que, a su vez, está conectado al puente de producción. [24]

- Equipos de superficie:

Una vez que se realiza el diseño de las instalaciones del subsuelo, se procede a dimensionar los equipos superficiales que puedan proveer la energía necesaria para hacer funcionar el sistema de bombeo. Para ello, las instalaciones superficiales deben poder soportar las cargas axiales de todo el equipo de fondo, proveer el torque necesario en el vástago, rotar a la velocidad requerida y ser capaces de prevenir la fuga de los fluidos producidos, todo esto teniendo en cuenta la seguridad del personal operativo. [13] [24]

- **Cabezal de rotación:** es el mecanismo encargado de soportar todo el peso de los equipos de subsuelo y los efectos producidos por el funcionamiento de la bomba, además, tiene que ser capaz de poder aguantar el efecto del mecanismo de frenado del sistema y de poder transmitir el movimiento a la sarta de varillas. [24]

- **Motor eléctrico:** es el equipo encargado de transmitir la energía necesaria a las instalaciones superficiales, para poder desempeñar las actividades de bombeo. Este motor puede ser eléctrico o de combustión. [13] [24]

- **Moto variador mecánico:** consiste en un conjunto de correas y poleas de distintos tamaños y sirven como acople entre el motor y la caja reductora. Gracias a este equipo no es necesario detener el motor para poder realizar el cambio de velocidad

de rotación de la bomba, en estos casos, las velocidades de rotación pueden variarse desde 50 hasta 400 RPM. [13]

- **Motorreductores:** este dispositivo, es el encargado de transmitir la potencia generada por el motor eléctrico (o de combustión) hacia el rotor de la bomba, manteniendo las condiciones de operación normal de estos elementos. En condiciones de operación normal, el motor opera a una frecuencia fija de aproximadamente 1800 RPM, mientras que la bomba, opera en un rango que va desde 40 a 400 RPM. Este cambio en la velocidad angular se logra gracias al motorreductor, que, adicionalmente al cambio de velocidad angular, suministra el torque necesario para la operación normal de la bomba. [13]

- **Variadores de frecuencia:** este equipo es necesario para que el sistema pueda realizar los cambios de velocidades en poco tiempo y sin la necesidad de modificar los equipos mecánicamente. El variador de frecuencia es un dispositivo electrónico que regula la corriente consumida por el motor, cambiando según convenga la frecuencia y el voltaje. Cuando la frecuencia es alterada, también ocurre un cambio en la velocidad de rotación, dado que ambos parámetros son proporcionales. Adicionalmente, el variador de frecuencia puede ajustar el torque durante las operaciones de rampeo, así como también reducir los picos de corriente que pudiesen afectar el sistema. [13]

- **Sistema de Frenado:** cuando el sistema de producción BCP se encuentra en operación normal, se acumulan grandes cantidades de energía en forma de torsión en la sarta de varillas. Si el sistema se detiene repentinamente, la energía acumulada se libera haciendo girar la sarta de cabillas en sentido inverso (a esto se le llama Back-Spin). Cuando esto sucede, las velocidades de rotación alcanzadas son muy altas, y podrían causar daños severos a los equipos o al operador. Para evitar esta situación, el cabezal de rotación cuenta con un sistema de frenado, que mantiene controlada la velocidad angular del *backspin*. [24] Los sistemas de frenado más comunes son:

- ✓ **Freno de accionamiento por fricción:** Este sistema de frenado está compuesto por un disco y pastillas de fricción, las cuales se accionan cuando

el disco comienza a girar en sentido contrario (*back-spin*). Este tipo de frenos se instala en el exterior del cabezal. [24]

- ✓ **Freno de accionamiento hidráulico:** a diferencia del sistema de frenado anterior, éste está integrado al cabezal y consta de un plato rotatorio que gira libremente durante el funcionamiento de la bomba. Cuando el *BackSpin* ocurre, se activa un sistema hidráulico que ofrece resistencia al movimiento inverso del plato rotatorio, lo que reduce la velocidad angular y, por lo tanto, disipa un poco la energía acumulada. [24]

Los elementos de las BCP que se mencionan en este capítulo son los básicos para un correcto funcionamiento. Existen otros elementos y componentes adicionales que mejoran o amplían el funcionamiento de dichas bombas.

En la siguiente sección se hablará de las aplicaciones de las BCP en la industria petrolera.

2.3.2.4.- Aplicaciones

Debido a su simplicidad en comparación con otros métodos de producción, además de sus bajos costos de inversión inicial y mantenimiento, el bombeo por cavidades progresivas se podría usar como la primera opción a considerar al momento de seleccionar un sistema de levantamiento artificial. [24]

Gracias al desarrollo de mejoras en las capacidades, las presiones de trabajo y en nuevos tipos de elastómero, las BCP pueden ser usadas en una gran variedad de pozos: pozos de crudos pesados y extrapesados con altos cortes de arena, crudos medianos (con limitaciones en contenido de H₂S), petróleos livianos (con limitaciones en contenido de aromáticos), pozos con altos cortes de agua, con bajas a medianas tasas de producción, pozos verticales, altamente desviados y horizontales, pozos con gas libre, entre otras. [24]

Sin importar la aplicación que se le dé a una BCP todas presentan las mismas características operativas, sin embargo, es necesario conocer también la clasificación de las bombas BCP las cuales se explican más adelante. Estos dos parámetros son fundamentales para la selección del tipo de bomba óptimo para un pozo.

2.3.2.5.- Características operativas de las BCP

Las bombas de cavidades progresivas tienen dos características principales: su desplazamiento volumétrico o caudal, y su altura de descarga o Head.

- Desplazamiento volumétrico o caudal:

El desplazamiento volumétrico representa el volumen hipotético de fluido que la bomba desplaza en un periodo de tiempo. Normalmente, los fabricantes representan esta cantidad en volumen de fluido por unidad de tiempo (bbl/día o m³/día) a determinadas condiciones de operación (normalmente a 100. o 500 RPM. y 0 head).
[13]

- Altura de descarga o head

La altura de descarga representa la capacidad de la bomba para poder vencer la presión hidrostática que se genera al transportar los fluidos desde la salida de la bomba hasta las instalaciones de superficie. Los fabricantes expresan esta cantidad como presión directamente (de acuerdo con el sistema de medición que mejor se adapte, unidades de campo o unidades métricas) o como altura de fluido (mts de columna de agua o pies de columna de agua) Es importante destacar que la altura de descarga es proporcional a la cantidad de etapas de la bomba. [13]

Como se había mencionado con anterioridad la clasificación de las BCP es otro parámetro fundamental para su selección, por lo tanto, en la siguiente sección se estudia la clasificación de las bombas BCP de acuerdo su uso, a su ensamblaje, según su funcionamiento y según su geometría.

2.3.2.6.- Clasificación de las BCP

De manera general, las bombas de cavidades progresivas se pueden clasificar en dos grandes grupos:

- Bombas industriales:

Son aquellas bombas (horizontales) que están destinadas al uso en diversas industrias (excepto la petrolera) como la industria alimenticia, la agropecuaria, en plantas de tratamiento de agua, entre otras. ^[13] ^[26]

- Bombas para la industria petrolera:

Son aquellas bombas que están especialmente diseñadas para su uso en la industria petrolera. Se usan para la extracción de hidrocarburos, el transporte de crudo, para la recuperación de lodos de perforación y para la transferencia de fluidos ácidos. ^[13] ^[26]

- Clasificación de las BCP de acuerdo con su ensamblaje

Las BCP se pueden clasificar de acuerdo con su ensamblaje en dos tipos, bombas tubulares y bombas de tipo insertable.

- Bombas tubulares

En este tipo de bombas, el estator y el rotor son elementos separados e independientes. Al momento de la instalación, el estator se baja con la tubería de producción junto a todos los accesorios necesarios (niple de paro, anclas de torque, anclas de gas, niples X, entre otros). Luego, una vez instalado el estator, el rotor se conecta y se baja mediante la sarta de cabillas. A diferencia de las bombas de tipo insertable, éstas pueden desplazar una mayor cantidad de fluido ya que el diámetro de la bomba es mayor. La desventaja que presenta es que, al momento de reemplazar la bomba, hay que recuperar toda la tubería de producción, por ende, se consume tiempo y recursos. ^[13] ^[27]

- Bombas tipo insertables:

A diferencia del tipo tubular, en este tipo de bombas, aunque el rotor y el estator son elementos separados se ensamblan en superficie, con el fin de bajar un solo conjunto, usando la sarta de cabillas, hasta que se ensambla sobre la zapata o sobre un niple de asentamiento. La ventaja que presenta sobre el tipo tubular es que para su reemplazo o extracción solo se sube la sarta de cabillas, ahorrando costos y tiempo, y con ello, producción diferida. La desventaja radica en que, al ir instalada dentro de la tubería de producción, su capacidad volumétrica está limitada, ya que su diámetro externo tiene que ser menor que el diámetro interno de la tubería de producción. ^[13] ^[27]

- Clasificación de las BCP de acuerdo con su geometría

Las BCP se pueden clasificar de acuerdo con su geometría en dos tipos, bombas de geometría simple o lobular y bombas multilobulares.

- Bombas de geometría simple o lobular

Son aquellas bombas en las que el rotor solo posee un lóbulo, mientras que el elastómero del estator es de dos lóbulos (relación 1:2). ^{[13] [24]}

- Bombas multilobulares

Son aquellas bombas en las que el rotor posee dos o más lóbulos, mientras que el estator posee un lóbulo más que el rotor (relación $n: n+1$). En general, estas bombas pueden entregar mayor cantidad de fluidos en comparación con bombas de geometría simple con características similares. ^{[13] [24]}

- Configuraciones especiales de BCP

Las BCP se pueden clasificar de acuerdo con su sistema de funcionamiento en dos tipos, *charge pumps* y BCP con sistema eléctrico de manejo en fondo (*Electric Downhole Drive PCP Systems*).

- *Charge Pumps*

Este tipo de bombas están compuestas por dos bombas de cavidades progresivas de diferentes capacidades volumétricas y capacidades de levantamiento colocadas en serie, separadas por una junta (*pup joint*). De estas bombas, la inferior es de mayor desplazamiento, pero de menor capacidad de levantamiento, mientras que la superior (la principal) posee un menor desplazamiento, pero una mayor capacidad de levantamiento. Esta configuración se usa mayormente en pozos con alto contenido de gas, ya que ayuda a elevar la eficiencia volumétrica de la bomba principal. Esto se logra

al comprimir el fluido producido usando la bomba de mayor desplazamiento, antes que dicho fluido entre a la bomba superior, lo que origina una presión de fondo mayor y ayuda al enfriamiento del fluido de la bomba primaria, lo que prolonga la vida útil de la misma. [28] [29]

- BCP con sistema eléctrico de manejo en fondo (*Electric Downhole Drive PCP Systems*)

En este sistema, la potencia que se le trasmite a la bomba de cavidades progresivas proviene de un motor eléctrico sumergido, lo que elimina la necesidad de la sarta de barras de bombeo, lo que las hace ideales para pozos desviados y terminaciones horizontales. Adicionalmente, a diferencia de las bombas regulares, este tipo de bombas requieren de los equipos usuales para las bombas electrosumergibles, como transformadores, variadores de frecuencia, caja de ventilación, entre otros. [28] [30]

A pesar de que existen diferentes tipos de clasificación de las bombas BCP, el proceso de diagnóstico y optimización es muy similar para todas.

2.3.2.7.- Diagnóstico y optimización

Cuando el sistema de levantamiento se coloca en marcha (sobre todo si es la primera vez que el pozo usa una BCP), es sumamente importante que se ajuste la velocidad de la bomba a una velocidad baja, de esta manera, los parámetros de producción como presión de cabezal, nivel dinámico de fluido, entre otros, se van estabilizando de manera progresiva. [13]

Cabe destacar que, en esta etapa inicial, el sistema de bombeo se va adaptando a las condiciones de operación normal, tales como, el contacto del elastómero con el tipo de fluido, el aumento de temperatura en la bomba, la elongación y torsión de las cabillas, el espaciamiento, entre otros. Por esa razón, es de suma importancia el

seguimiento cuidadoso de los parámetros de operación en esta etapa, pues ésta es la más crítica, en cuanto a la cantidad de fallas se refiere. [13]

Cada vez que el sistema se estabilice a una nueva velocidad de rotación, es necesario realizar mediciones para luego poder compararlas con los cálculos teóricos arrojados por el simulador. Algunas de las variables que se miden son:

- ✓ Tasa de producción.
- ✓ Nivel de fluido
- ✓ Relación gas-petróleo
- ✓ Porcentaje de agua y sedimentos
- ✓ Presión de cabezal
- ✓ Parámetros de operación de la bomba (RPM, frecuencia, torque, potencia, etc.) [13]

En el Manual del *Workshop International* ‘Bombeo de Cavidad Progresiva’, dictado por el Ing. Nelvy Chacin (2003), se explica un procedimiento manual para la optimización de un pozo con sistema de bombeo BCP (Anexo 2).

Actualmente, existen unos dispositivos denominados *Data Logger*, que se usan para automatizar el monitoreo en tiempo real de las variables críticas de operación. Algunos incluso tienen incorporados algoritmos avanzados para poder realizar algunas tareas complejas (por ejemplo, estimar niveles de fluido o analizar comportamientos históricos) y en base a eso, tomar decisiones que apunten a la optimización del pozo. [13]

2.3.2.8.- Mantenimiento de los equipos

Una BCP requiere de poco mantenimiento, en comparación con otros métodos de levantamiento, como se había mencionado con anterioridad. Los equipos de subsuelo no requieren casi mantenimiento, pues en caso de que la bomba no cumpla con los requerimientos diseñados, se procede a sustituirla. En estos casos, la bomba

que se extrae se somete a una inspección en un taller con un banco de pruebas para determinar si aún se puede utilizar de manera total o parcial en otro pozo cuyos requerimientos de caudal y/o de head sean menores. ^[13]

En lo que se refiere a los equipos de superficie, la mayor parte del mantenimiento que requieren, son los relacionados con la lubricación de éstos. Es importante que los mecanismos móviles se mantengan debidamente engrasados y que los niveles de aceite de lubricación de los rodamientos del cabezal y la caja reductora se mantengan en sus niveles óptimos de acuerdo con el fabricante. Hay que destacar que la grasa y los aceites requeridos para el mantenimiento deben poseer las propiedades necesarias de acuerdo con el ambiente en donde las máquinas estén operando. Para este fin se debe consultar al fabricante. ^[13]

El mantenimiento regular de los equipos es de suma importancia, ya que gracias a ellos se pueden evitar un gran número de problemas. En la figura 14, se observan las fallas más comunes que afectan el sistema BCP por la falta de mantenimiento, donde la falla más común es la avería en las poleas del cabezal, sin embargo, esto puede variar de acuerdo con las condiciones del sistema pozo yacimiento y el tipo de fluido en producción: ^[13]

Trabajos realizados en los equipos de superficie

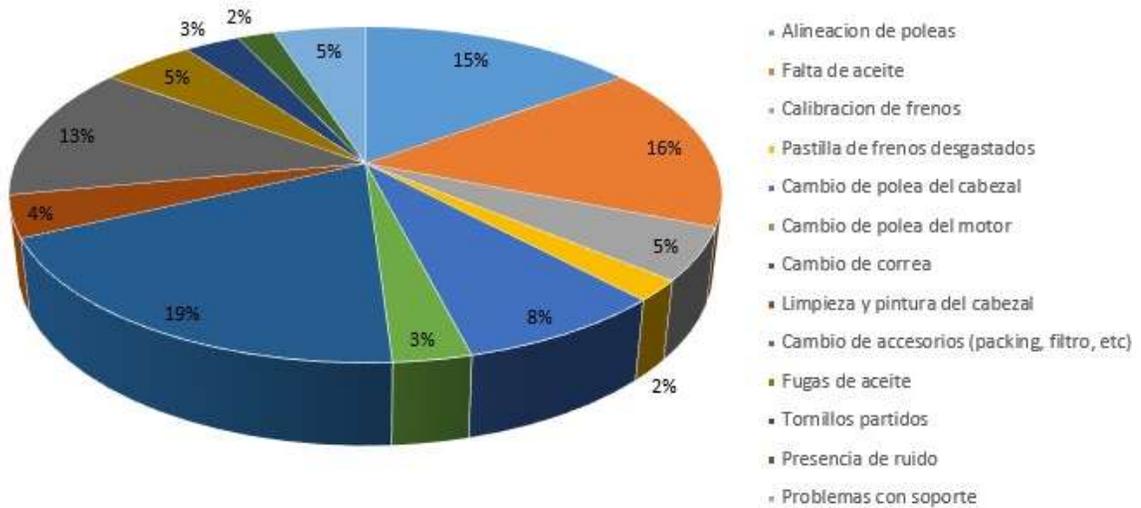


Figura 14 Gráfico de trabajos realizados en los equipos de superficie

Fuente: Frecuencia de fallas de los equipos integrales de superficie por falta de mantenimiento, pág. 118 (Modificado) Nelvy Chacón. Bombeo de Cavidad Progresiva

Las bombas BCP presentan diferentes problemas asociados a fallas eléctricas, fallas mecánicas, entre otras. Con el tiempo han sido propuestas diversas causas y posibles soluciones que pueden servir de referencia para mejorar el rendimiento de una BCP y así poder determinar de forma más rápida una falla en campo (Anexo 3).

En las secciones anteriores se ha hablado de las definiciones básicas de yacimiento y los sistemas de producción, sin embargo, se deben estudiar todos estos parámetros desde el punto de vista económico y de negocio.

2.4.- El rol de la producción de petróleo en la demanda energética mundial

La energía juega un papel fundamental en la sociedad moderna, la cual, ha experimentado profundos cambios en su estilo de vida, pasando de una economía de subsistencia a una economía industrializada y con un gran número de servicios. La demanda energética mundial ha incrementado en más de un 40% desde el año 1993, de acuerdo con el *BP Statistical Review 2019*. En la figura 15 se muestra el gráfico del consumo energético mundial, en la cual se observa cómo ha aumentado la demanda energética de acuerdo con los diferentes recursos energéticos, como el petróleo, carbón, gas natural, entre otros. Como consecuencia de esto las empresas petroleras han optado por aumentar la producción y optimizar los procesos constantemente. [31] [32]

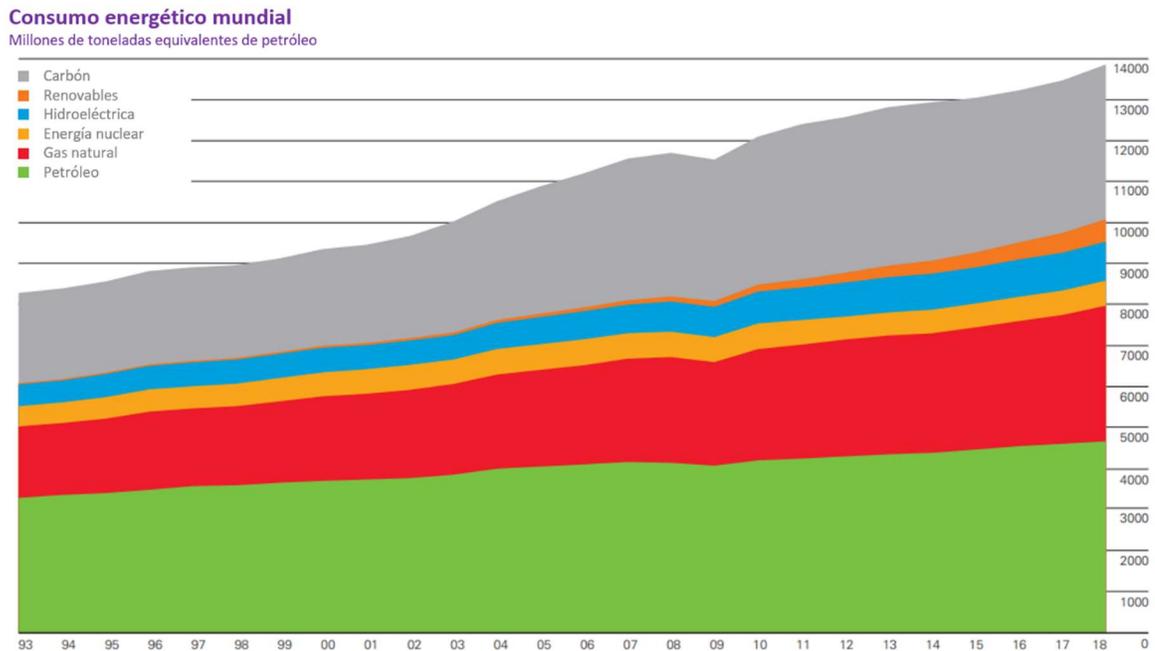


Figura 15: Gráfica del consumo energético mundial desde el año 1993 hasta el 2018 (Modificado).

Fuente: *BP Statistical Review 2019*.

<https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2019-full-report.pdf>

La mayor parte de las reservas de petróleo mundial corresponden a petróleo pesado, extrapesado, arenas petrolíferas o bitumen de alta viscosidad, como se muestra en la figura 16, donde el 70% de las reservas corresponden a este tipo de petróleo. [33]

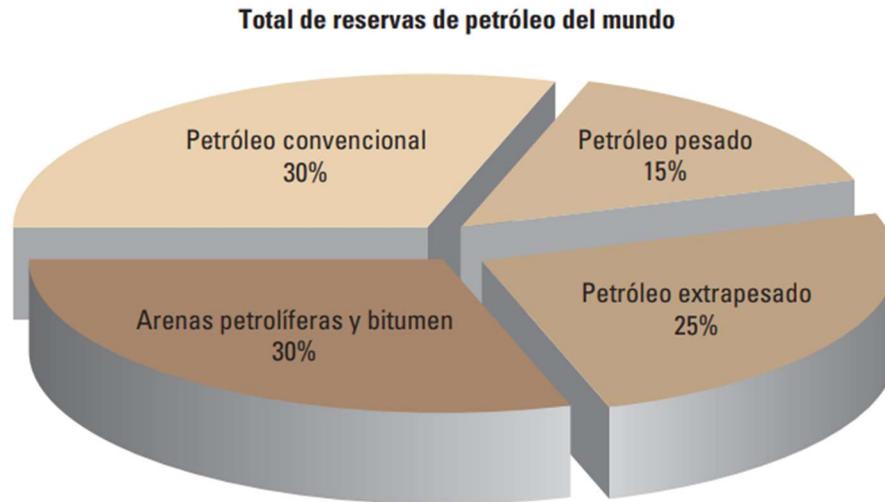


Figura 16: Total de reservas de petróleo del mundo

*Fuente: La importancia del petróleo Schlumberger (Documento PDF)
<https://www.slb.com/-/media/files/oilfield-review/heavy-oil-3-spanish>*

La gran demanda de petróleo en los últimos años ha generado una declinación en la producción de petróleo en yacimientos convencionales, como consecuencia las empresas petroleras se han visto obligadas a explotar los yacimientos de petróleo pesado. Es importante tomar en cuenta que mientras más pesado es el petróleo, su valor comercial es menor, y por su alta viscosidad son más difíciles de producir y refinar. Los avances tecnológicos han hecho posible que producir este tipo de petróleo sea rentable, aunque las ganancias son menores. Por esta razón, las empresas siempre buscan optimizar los procesos de producción y reducir los costos. ^[33]

En la actualidad la ciencia de datos es una herramienta computacional muy efectiva y poderosa que les permite a los ingenieros realizar análisis de datos en poco tiempo; con la finalidad de generar soluciones efectivas, optimizar producción y una rápida toma de decisiones. En la siguiente sección se presentan varios ejemplos de cómo la ciencia de datos ha sido utilizada en las ramas de ingeniería y en la industria petrolera con resultados positivos.

2.5.- Antecedentes del uso de *Data science*

Data Science o también conocido como la ciencia de los datos es un conjunto de disciplinas preexistentes; este término ha sido utilizado desde principios de los años sesenta cuando John Tukey escribió “*The Future of Data Analysis*”, generando una relación entre el análisis y la estadística. Con los años, el desarrollo de esta ciencia ha ido creciendo de manera exponencial. En el año 2001, el científico Williams Cleveland escribió un artículo titulado “*Data Science: An Action Plan for Expanding the Technical Areas of the field of Statistics*”, en el cual planteó la ciencia de datos como una disciplina independiente y la dividió en seis importantes ramas: investigación multidisciplinaria, modelos y metodologías aplicadas a datos, computación aplicada al análisis de datos, pedagogía, evaluación de herramientas y teoría. [34]

A principios del siglo XXI, la ciencia de datos estaba presente en compañías como IBM, las cuales demostraron la relación existente entre la ciencia de datos y el aprendizaje automatizado. Grandes compañías como Facebook, Amazon y Google han usado la ciencia de datos para mejorar sus ventas, optimizar los procesos de marketing y negocios, diseñar productos inteligentes y predecir el comportamiento de los usuarios y clientes. De igual forma laboratorios y clínicas han usado este tipo de metodologías para controlar y predecir la evaluación de epidemias y enfermedades. [35]

El uso de *Data Science* en la actualidad ha desencadenado grandes avances y resultados favorables para la sociedad, donde la ingeniería ha sido pionera de estos descubrimientos. En el área petrolera, compañías como Schlumberger, Chevron, Devon Energy, entre otras, han aplicado la ciencia de datos con el fin de mejorar los resultados y optimizar los procesos dentro de sus empresas. Este tipo de metodologías se han aplicado para la optimización de producción, localización de pozos, cotejo histórico, optimización de perforación, correlaciones PVT y predicciones de presión y tasas de producción. [36]

Devon Energy desde el año 2012 comenzó a invertir en tecnología para la aplicación de *Data Science*. La compañía desarrolló un departamento multidisciplinario basado en la aplicación de inteligencia artificial y aprendizaje

automatizado, con la finalidad de generar algoritmos sofisticados que pudiesen ser aplicados en distintos departamentos como exploración, perforación, completación y producción de yacimientos. En el año 2016 la compañía tenía la capacidad de analizar datos en pocos minutos y realizar una rápida toma de decisiones basado en las aplicaciones y algoritmos diseñados. Para la fecha la compañía podía analizar procesos como: pruebas piloto de pozos, productividad de un pozo basado en el espaciamiento entre los mismos, tiempos esperados de perforación, predicción de geo amenazas en distintas zonas, selección del ensamblaje de fondo adecuado, análisis de riesgo del fracturamiento hidráulico en tiempo real, optimización y mantenimiento preventivo de los mecanismos de levantamiento artificial en los pozos. Con este tipo de metodologías Devon logró aumentar su productividad en pozos horizontales un 250%, y reducir los costos de perforación, completación y operación un 40 %. [37]

En el año 2015, trabajadores de la compañía Anadarko Petroleum en conjunto de la universidad de Standford, plantearon una metodología llamada “*Functional Data Analysis*” (FDA) basada en estadística y aprendizaje automatizado, en vista de la dificultad de aplicar técnicas de modelado común o simulación tradicional en yacimientos no convencionales para la predicción de la producción. Esta metodología modela el yacimiento como series de tiempo, compuestas por distintas funciones analíticas ponderadas, las cuales permiten predecir la producción del yacimiento de forma no paramétrica, ni confinado a un modelo de declinación previamente formulado[38]. En el año 2016, la compañía Schlumberger logró implementar el aprendizaje automatizado con redes neuronales para la predicción de la producción en un yacimiento no convencional en Argentina. El entrenamiento de esta red se basó en el historial de producción, mapas geológicos, datos de presión y restricciones operacionales con el fin de estudiar el comportamiento de los pozos ya perforados y la predicción de pozos nuevos. Cada uno de estos parámetros de entrada eran necesarios para disminuir el error o incertidumbre de la predicción, el historial de producción fue utilizado para estudiar la respuesta del pozo a diferentes restricciones operacionales como: la presión de cabezal o el tamaño de *tubing*, los datos geológicos, les permitieron obtener información acerca de la productividad del pozo respecto al yacimiento,

espesores, porosidad promedio, entre otros, y por último la ubicación de los pozos, les permitió evaluar el efecto de interferencia entre los pozos cercanos. En la figura 17, se muestra el resultado obtenido al aplicar esta metodología, los puntos color azul representan la producción de petróleo real, mientras los puntos en color naranja representan la predicción obtenida del algoritmo de entrenamiento y los puntos verdes la validación del modelo. El resultado representado en la gráfica demuestra la factibilidad de aplicar el aprendizaje automatizado a la industria petrolera. [39]

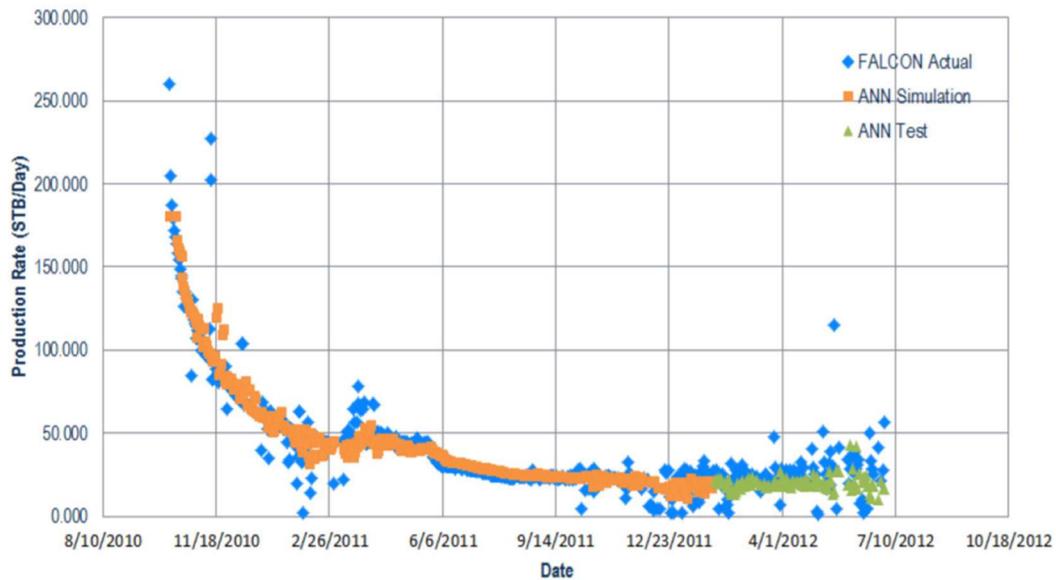


Figura 17: Comparación entre la producción real y los valores obtenidos por la red neuronal

Fuente: Paper SPE 180984-MS, Schlumberger pág. 6

El aprendizaje automatizado permite entender mejor los comportamientos anormales en el sistema de producción. En el año 2015, ingenieros de la compañía Petrobras desarrollaron una metodología para realizar mantenimiento preventivo a los equipos de producción. Estos se basaron en un sistema de monitoreo en los procesos de deterioro de los equipos mediante el uso de datos adquiridos durante su funcionamiento, con la finalidad de pronosticar, detectar, clasificar y predecir fallas críticas.

La optimización de los costos de mantenimiento es una de las principales preocupaciones de los operadores, por lo tanto, la capacidad de predecir fallas críticas es un factor clave. Este procedimiento consta de seis etapas básicas: el análisis del funcionamiento de la máquina, análisis de los componentes, modelado de los síntomas, descripción de la falla, procesamiento y reconocimiento de las fallas a través de algoritmos diseñados, y por último el diagnóstico y la prognosis para la toma de decisiones.^[40]

En estos antecedentes se puede observar como la ciencia de datos ha funcionado como una herramienta muy poderosa que puede brindar grandes soluciones para la industria. En las siguientes secciones se definen conceptos básicos de estadística relacionados con la ciencia de datos, metodologías y disciplinas de *data science*, las cuales son fundamentales para este TEG.

2.6.- Estadística

La estadística se define como una disciplina científica la cual se encarga de la obtención, orden y análisis de un conjunto de datos, con la finalidad de obtener explicaciones y predicciones de fenómenos observados. Se basa en métodos, procedimientos y ecuaciones que permiten recolectar información, analizarla y extraer las conclusiones necesarias. Entre los parámetros y metodologías importantes de la estadística que se utilizarán en este TEG tenemos la normalización, la estandarización y por último la significancia estadística (*p-values*).^[41]

2.6.1.- Normalización

La normalización es un proceso que te permite comprimir o extender los valores de una variable para que todos estén en un rango definido. Este proceso se calcula restando el valor mínimo del parámetro en estudio de cada uno de los valores y dividirlo

entre la diferencia existente del valor máximo y el valor mínimo de dicho parámetro, como se muestra en la siguiente ecuación:

$$x = \frac{x - \min(x)}{\max(x) - \min(x)} \text{ Ec. 8}$$

2.6.2.- Estandarización

Se define como el proceso de poner diferentes variables en la misma escala, con la finalidad de poder comparar sus valores. Por lo general, para estandarizar las variables, se calcula la media y la desviación estándar de la variable y luego, para cada valor observado de la variable, se le resta la media y se divide por la desviación estándar, como se muestra en la siguiente ecuación:

$$x = \frac{x - \text{media}(x)}{\text{desviacion estandar}(x)} \text{ Ec. 9}$$

2.6.3.- Significancia estadística (*p-values*)

La significancia estadística se basa en determinar cuándo un resultado es estadísticamente significativo y que este resultado no se debió al azar. Se puede definir como la probabilidad que tiene una variable en afectar o no la hipótesis planteada.

La significancia estadística se puede calcular con el valor p (*p-value*), el cual se define como el valor de la probabilidad que la correlación entre dos variables sea estadísticamente significativa. Por lo general se utiliza un nivel de significancia de 0,05 lo que significa que estamos 95 % seguros de que la correlación entre las variables es significativa. Basados en este concepto podemos decir que:

- Si el valor de p es menor a 0.001 ($p < 0.001$) se dice que existe una fuerte evidencia de que la correlación es significativa

- Si el valor de p es menor a 0.05, pero, mayor a 0.001 ($0.001 < p < 0.05$) se dice que existe evidencia moderada de que la correlación es significativa
- Si el valor de p es mayor a 0.05, pero, menor a 0.1 ($0.05 < p < 0.1$) existe evidencia débil de que la correlación es significativa
- Si el valor de p es mayor a 0.1 ($p > 0.1$) indica que no hay evidencia que la correlación sea significativa

Por lo tanto, el valor de p puede indicar que tan importante es un parámetro para la predicción.

Estos conceptos y procesos estadísticos básicos son fundamentales para la ciencia de datos, la cual se define a continuación.

2.7.- Data science

Data science (o ciencia de los datos) se conoce como un área de estudio multidisciplinaria, en el que se integran ciencias básicas como la estadística, la matemática y la computación, con la finalidad de analizar un conjunto de datos y extraer información valiosa tomando en cuenta distintos modelos e hipótesis, que permitan transformar la información en un valor tangible empresarial. ^{[42][43]}

La ciencia de datos emplea complejas metodologías y disciplinas encargadas de la limpieza, preparación y análisis de los datos. Para llevar a cabo un estudio de *Data science*, existen dos metodologías básicas: SEMMA y CRISP-DM que se estudian a profundidad en este capítulo. Sin embargo, para la aplicación de estas metodologías es necesaria la integración de distintas disciplinas que permitan realizar el proceso de manera más eficiente.

2.7.1.- Metodologías de data science

En el año 1989, se empleó por primera vez el término KDD (*Knowledge Discovery in Databases*) para hacer referencia al concepto de hallar información

relevante en un conjunto de datos. Si bien este término describe todas las actividades y métodos usados para extraer información que se considera útil, lo hace de manera muy amplia y poco específica. Años después (hacia finales de los años 90), se intentó normalizar este proceso de análisis, lo que trajo como resultado, el surgimiento de dos metodologías de trabajo que son las más usadas actualmente en los proyectos de *data science*: SEMMA (*Sample, Explore, Modify, Model, Assess*) y CRISP-DM (*Cross-Industry Standard Process for Data Mining*). ^{[44][45]}

2.7.1.1.- Metodología SEMMA

Es una metodología de trabajo para llevar a cabo un proyecto de *Data Science*, la cual fue desarrollado por el *SAS Institute (Statistical Analysis System Institute)*. Esta metodología consiste en 5 etapas, ^{[44][46]} que son:

- *Sample* (muestreo): en esta primera etapa se selecciona una muestra de los datos que se van a analizar. La muestra tiene que ser lo suficientemente grande para extraer información importante, pero lo suficientemente pequeña para hacer este proceso de manera eficiente. ^{[44][46]}
- *Explore* (exploración): en esta segunda etapa se exploran los datos seleccionados para tener una idea preliminar de las relaciones y tendencias existentes entre las variables, así como también la búsqueda de anomalías que pudiesen afectar los resultados. ^{[44][46]}
- *Modify* (modificación): en la tercera etapa se manipulan los datos y se crean nuevas variables con el fin de preparar y enfocar el proceso de modelado. ^{[44][46]}
- *Model* (Modelado): en esta etapa se aplican varias técnicas de modelado usando softwares especializados, con la finalidad de buscar combinaciones en los datos que produzcan una respuesta deseada y confiable. ^{[44][46]}
- *Assess* (Evaluación): en esta última etapa se evalúa el modelo resultante para poder determinar su utilidad y confiabilidad, y finalmente estimar su desempeño. ^{[44][46]}

2.7.1.2.- Metodología CRISP-DM

La metodología CRISP-DM fue diseñada en 1996. Su desarrollo fue liderado por 5 compañías (*Integral Solutions Ltd., Teradata, Daimler AG, NCR Corporation y OHRA*). A diferencia de la metodología SEMMA, CRISP-DM toma en cuenta la aplicación al entorno del negocio, y considera, además, la implementación y mantenimiento del modelo, por eso, de acuerdo con encuestas recientes se demostró que esta metodología es la más aceptada y usada en la actualidad.

Esta metodología, consta de 6 etapas: ^{[45][47]}

- *Business Understanding* (Comprensión de negocio): en esta etapa se definen las necesidades del cliente y se estudian los objetivos y requerimientos del proyecto desde el punto de vista del negocio. Además, se determinan los factores que pudiesen afectar el resultado del proyecto, así como diseñar un plan preliminar para poder lograr los objetivos. ^{[45][47]}
- *Data Understanding* (estudio y comprensión de los datos): abarca las actividades relacionadas con la adquisición de datos y su estudio inicial a fin de familiarizarse con ellos. Busca identificar los problemas de calidad que pudiesen tener los datos, y hacer un análisis preliminar, para poder descubrir posibles tendencias o patrones que sirvan para formular las hipótesis iniciales. ^{[45][47]}
- *Data Preparation* (Preparación de datos): durante esta etapa, se realizan todas las actividades necesarias para obtener el conjunto de datos que se va a usar para el modelo final. Primero, se seleccionan los datos a utilizar, tomando en cuenta las fuentes de los datos, los tipos de datos, su estructura, etc. Luego, se realiza una limpieza de datos, estimando o eliminando valores faltantes o incongruentes. Una vez que los datos estén sin errores, se determina si se requieren datos adicionales, para lo cual, se pueden generar nuevos campos a partir de la manipulación de valores existentes. Adicionalmente se realizan procesos de normalización y estandarización para disminuir el tiempo de cómputo. ^{[45][47]}

- *Modeling* (Modelado): una vez que los datos están preparados, se aplica la técnica de modelado que se va a usar para desarrollar el proyecto (por ejemplo, árbol de decisiones o redes neuronales). Una vez se haya seleccionado el método a usar, se procede a diseñar una prueba de validación con sus respectivos parámetros, con la finalidad de evaluar la calidad del modelo. Para poder llevar a cabo la prueba, es necesario tomar una parte de los datos para tal fin. Finalmente, se construye el modelo usando los datos que se seleccionaron para el entrenamiento y se calibran algunos parámetros para un resultado óptimo. Los resultados deben interpretarse de acuerdo con los objetivos del negocio. ^{[45][47]}
- *Evaluation* (Evaluación): en esta etapa del proceso, el modelo generado se somete a una evaluación más exhaustiva, haciendo un recuento de los pasos realizados para generarlo, y determinar si los objetivos de negocio planteados al principio fueron alcanzados. Adicionalmente, se realiza una revisión de los datos, para cerciorarse que no se haya pasado por alto algún factor importante. Finalmente, tomando en cuenta los resultados se toma la decisión de hacer una nueva iteración del proyecto o pasar a la fase de despliegue. ^{[45][47]}
- *Deployment* (Despliegue): Esta es la última etapa de esta metodología. Una vez que el modelo ya ha sido evaluado y ha cumplido los objetivos planteados en el primer paso, se diseña una estrategia para el despliegue del modelo en el negocio. En esta etapa, los resultados del modelo generado se deben organizar y presentar de tal manera, que el cliente comprenda el análisis y que lo pueda implementar con éxito en su negocio. Dependiendo del tipo de negocio y las necesidades de éste, el despliegue puede abarcar algo sencillo como la presentación de un informe detallado con los resultados obtenidos, o algo más complejo, como la automatización del proceso de análisis de datos, a fin de realizarlo de manera periódica. ^{[45][47]}

A continuación, se presenta una tabla comparativa, en donde se muestra una equivalencia entre las etapas de las dos metodologías descritas anteriormente. En ella se observa que la metodología CRISP-DM es la más completa, y por ende, es la más usada y aceptada en la actualidad:

Tabla 2: Comparación entre las etapas de la metodología SEMMA y CRISP-DM

SEMMA	CRISP-DM
-	Business Understanding
Sample	Data Understanding
Explore	
Modify	Data Preparation
Model	Modeling
Assessment	Evaluation
-	Deployment

2.7.2.- Disciplinas asociadas a Data Science

Para llevar a cabo un estudio de *Data science*, es necesaria la integración de distintas disciplinas, como lo son: *data mining*, *big data*, *inteligencia artificial*, *machine Learning* y *Deep Learning*. En la figura 18, se muestra un diagrama en la cual se observa cómo se relacionan estas disciplinas con la ciencia de datos. [48]

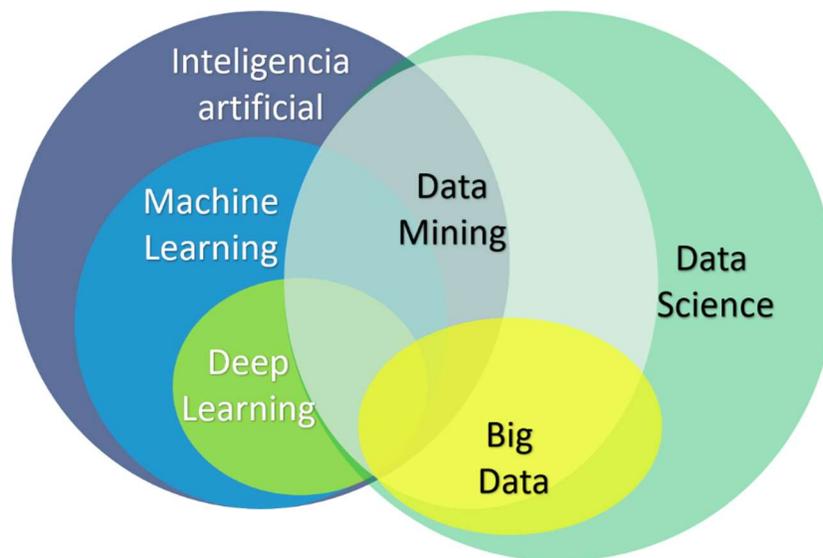


Figura 18: Disciplinas y metodologías asociadas a Data Science (Modificado)

Fuente: Manfre Diego. Clase de tópicos VIII, UCV (2020)

2.7.2.1.- Minería de datos (*data mining*)

Es el proceso básico que se encarga de recopilar información “cruda” (*raw data*) y extraer la información con mayor relevancia, es decir, procesarla con el fin de poder transformarla de manera apropiada en un conjunto de datos estructurados que permitan encontrar relaciones o tendencias entre los datos de forma rápida. [49][50]

2.7.2.2.- *Big data*

Se define como un conjunto de datos de gran volumen, complejidad y variabilidad que no pueden ser transportados, ordenados, manipulados o procesados con recursos convencionales, por lo general se caracterizan por ser datos no organizados, de múltiples fuentes y diferentes formatos. [51][52]

2.7.2.3.- Inteligencia artificial (*artificial intelligence*)

Este término hace referencia a las computadoras o máquinas que tienen la capacidad de imitar la inteligencia humana, es decir, que son programadas para aprender, razonar y autocorregirse, entre otras funciones cognitivas que son atribuidas al ser humano (como la resolución de problemas complejos). [53][54] [55][56] Se pueden clasificar en dos tipos:

- Inteligencia artificial débil (*Weak AI o Narrow AI*): es aquella ‘inteligencia’ que es diseñada para un solo propósito, por lo tanto, pueden realizar una tarea específica muy bien. Ejemplos de este tipo de inteligencia pueden ser los programas de reconocimiento facial o asistentes personales virtuales, como Siri (*Apple*) o Alexa (*Amazon*). [54]

- Inteligencia artificial general o fuerte (*Artificial General Intelligence/AGI o Strong AI*): es un sistema diseñado con habilidades cognitivas generalizadas. Pueden resolver problemas de manera inteligente. [54]

2.7.2.4.- Aprendizaje automatizado (*machine learning*)

Es una técnica dentro de la rama de inteligencia artificial que se encarga de proveer a las computadoras la habilidad de aprender a partir de un grupo de datos sin usar instrucciones explícitas para ello. Esta técnica se explica con mayor detalle en la sección 2.7 de este trabajo. ^[57]

2.7.2.5.- Aprendizaje profundo (*deep learning*)

Es una técnica de inteligencia artificial que se considera un subcampo más avanzado de *machine learning*. Esta técnica también usa los datos con la finalidad que la maquina aprenda, pero a diferencia de *Machine Learning*, su aprendizaje es a través de múltiples capas. Estas capas procesan la información extrayendo las características relevantes en cada una de las capas y estas características obtenidas se utilizan como información de entrada para las capas sucesivas, de esta manera, estos procesos pueden llegar a un resultado óptimo, muchas veces sin la intervención humana. ^{[54] [58][59]}

2.8.- Aprendizaje automatizado (*Machine learning*)

El aprendizaje automatizado (ML por sus siglas en inglés) es una técnica de inteligencia artificial aplicada a *Data Science*, que es utilizada para el análisis de datos y encontrar tendencias o patrones en los mismos, con la finalidad de generar conclusiones y ayudar a una rápida toma de decisiones. François Chollet, en su libro *Deep Learning with Python*, estipula que:

“Un sistema de Machine-Learning es entrenado en lugar de ser programado explícitamente. Se le presentan varios ejemplos relevantes para una tarea, y halla estructuras estadísticas en esos ejemplos que eventualmente le permiten al sistema proponer reglas para la automatización del proceso”.

En la programación clásica, el programador establece un conjunto de reglas y los datos de entrada (*input*) se procesan de acuerdo con estas reglas, finalmente el programa presenta unos resultados (*output*) de acuerdo con las reglas establecidas. En los algoritmos de *machine learning*, al programa se le presentan los datos de entrada (*input*) y sus correspondientes datos de salida (*output*), estos datos se procesan y el resultado son las reglas que rigen la relación entre los parámetros de entrada y los de salida. El objetivo de esto es que las reglas que el algoritmo encuentra puedan ser usadas para procesar nuevos datos de entrada y así producir nuevos datos de salida. En la figura 19, se muestra un esquema detallado donde se comparan ambas estrategias de programación estudiando los parámetros de entrada y de salida antes mencionados. [60]



Figura 19: Diferencia entre la programación clásica y machine learning.

Fuente: Chollet, François. *Deep Learning with Python*. 2018. Pág. 6

Actualmente existen varias metodologías de *machine learning*, pero las 3 más comunes son: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje reforzado.

2.8.1.- Aprendizaje supervisado

El aprendizaje supervisado es una de las metodologías más usadas actualmente. Estos algoritmos son ‘alimentados’ (o entrenados) con datos etiquetados, es decir, que para cada dato de entrada le corresponde un dato de salida (etiqueta o *label*). De esta manera, el algoritmo ‘aprende’ a identificar correctamente que etiqueta le corresponde

a cada parámetro de entrada (*input*). El objetivo es que el algoritmo modele reglas (calcular los coeficientes) que permitan etiquetar los datos de entrada, con la finalidad de poder estimar el resultado de aquellos datos que no formen parte de los datos de entrenamiento. [60]

2.8.2.- Aprendizaje no supervisado

En esta metodología, a diferencia del aprendizaje supervisado, el algoritmo se entrena con parámetros de entrada que no están etiquetados, en otras palabras, no se le da la respuesta correcta de cada dato de entrada. El objetivo de este tipo de algoritmos es hallar estructuras dentro de los *inputs*, para luego poder realizar tareas de clasificación con otros parámetros que no estén dentro de los datos de entrenamiento. [60]

2.8.3.- Aprendizaje reforzado (*Reinforcement learning*)

Esta metodología se ha popularizado recientemente, y es usado mayormente en las áreas de robótica, videojuegos y navegación. En el aprendizaje reforzado, existen 3 componentes principales: el agente (el que aprende o el que toma las decisiones), el ambiente (todo con lo que el agente pueda interactuar) y las acciones (lo que el agente pueda hacer). El agente recibe información del ambiente y aprende a tomar decisiones con el fin de maximizar las ganancias. El objetivo principal de este tipo de aprendizaje sería aprender a tomar decisiones con el fin de maximizar premios o ganancias en un periodo de tiempo. Actualmente solo ha tenido éxito en el campo de los videojuegos, pero se espera que se desarrolle lo suficiente para poder aplicarlo con éxito en el mundo real. [60]

2.8.4.- Metodologías de *machine learning*

En la actualidad son muchas las metodologías desarrolladas para el aprendizaje automatizado, sin embargo, este TEG se enfocará en la aplicación de seis de ellas, regresión lineal multivariable, regresión polinómica, *k-nearest neighbors*, *support vector machine*, árbol de decisiones y redes neuronales:

2.8.4.1.- Regresión lineal multivariable

Los análisis de regresión son un subcampo de *ML* supervisado, donde su propósito es establecer un modelo que permita relacionar una cierta cantidad de parámetros o características llamadas variables independientes con una variable objetivo, la cual recibe el nombre de variable dependiente. [61]

Por lo general, los problemas de regresión buscan respuestas cuantitativas, lo que permite realizar la predicción de un valor en particular.

Para entender la regresión lineal multivariable se debe partir de la teoría de regresión lineal simple. El modelo de regresión lineal simple es un método estadístico que permite estudiar la relación entre dos variables continuas cuantitativas. El algoritmo de regresión establecerá un modelo para ajustar la relación de dependencia entre una característica específica independiente (un valor de la variable independiente “x”) y el valor “resultado” correspondiente (un valor de la variable dependiente “y”). Matemáticamente, este tipo de regresión usa una función lineal para aproximar o predecir la variable dependiente. En la figura 20, se muestra la expresión matemática de este tipo de regresión y sus elementos, donde:

- y: variable dependiente o a predecir
- x: variable independiente
- a: pendiente o coeficiente determinado por el algoritmo
- b: constante determinada por el algoritmo, que representa el corte con el eje y.

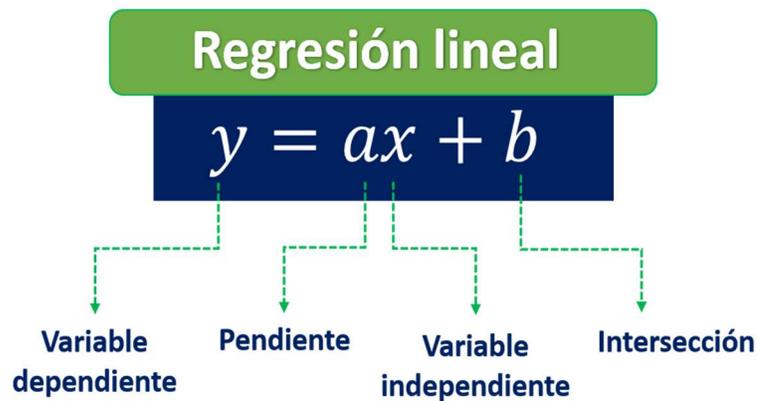


Figura 20: Algoritmo de regresión lineal (Modificado)

Fuente: <https://ligdigonzalez.com/algoritmo-regresion-lineal-simple-machine-learning/>

Esta relación se realiza estableciendo una línea arbitraria y calculando la distancia mínima de la recta a los puntos de datos correspondientes a los valores (x, y). En la figura 21, se muestra un ejemplo de una regresión lineal, donde se observa esta distancia mencionada, que se considera como “residuos” o “errores de la predicción” donde se representa en el eje x las RPM y en el eje y, la tasa de petróleo.

El algoritmo de regresión recalculará y moverá la recta con cada interacción, buscando aquella que mejor se ajuste a los puntos de datos, o, en otras palabras, la recta con el menor error. [61]

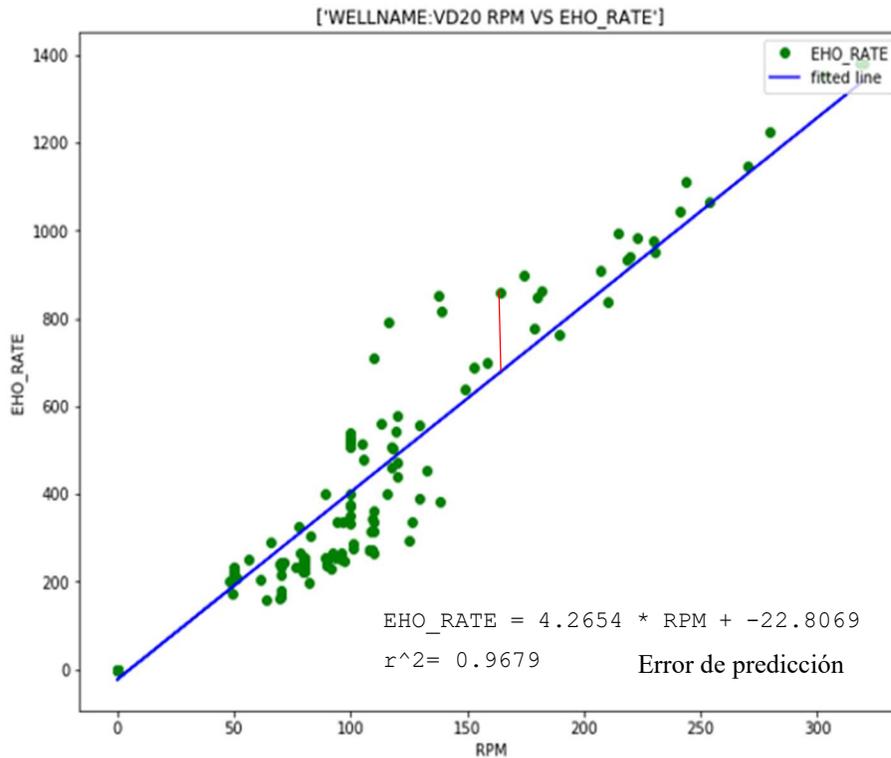


Figura 21: Ejemplo de Regresión lineal Simple

La regresión lineal múltiple es similar a la simple, la diferencia en ésta es que, en la múltiple se maneja más de una variable independiente que contribuyen a la variable dependiente, por lo tanto, se manejan múltiples coeficientes y su complejidad computacional es mayor. La ecuación matemática que define este estudio se ve representada en la figura 22, donde se muestra cada uno de los elementos que intervienen en la misma.

2.8.4.2.- Regresión polinómica

La regresión polinómica es considerada un caso especial del análisis de regresión lineal, en la cual, el modelo se extiende agregando predictores adicionales, que se obtienen al elevar cada una de las variables a una potencia. De acuerdo con Ligdi Gonzales (2017), especialista en *Machine Learning*, indica que:

“El método estándar para extender la regresión lineal a una relación no lineal entre las variables dependientes e independientes, es reemplazar el modelo lineal con una función polinomial”.

En la figura 23, se muestran las expresiones matemáticas que describen las distintas regresiones aplicadas a ambos casos, para una variable y para múltiples variables.

Regresión lineal simple

$$y = ax + b$$

Regresión polinómica simple

$$y = a_1x_1^2 + a_2x + b$$

Regresión lineal multivariable

$$y = a_1x_1 + a_2x_2 + \dots + a_nx_n + b$$

Regresión polinómica multivariable (2 variables) de grado 2

$$y = a_1x_1^2 + a_2x_1 + a_3x_2^2 + a_4x_2 + a_5x_1x_2 + b$$

Figura 23: Comparación entre las ecuaciones de regresión (Modificado)

Fuente: <https://ligdigonzalez.com/algorithmo-regresion-polinomial-machine-learning/>

Los términos polinomiales convierten el modelo de regresión lineal en una curva, sin embargo, el coeficiente de intersección calculado por el algoritmo no se ajusta a un elemento polinomial, es por esta razón que esta metodología de regresión

polinomial se considera un ajuste de la regresión lineal. Este tipo de regresiones presentan grandes problemas, debido a que a medida que aumenta el número de características y el grado del polinomio, aumenta la complejidad de la ecuación, lo que hace difícil los cálculos y el manejo del algoritmo. Por otro lado, en este tipo de regresiones a medida que aumentan su grado tienden a ajustarse excesivamente lo que puede generar problemas al momento de realizar una predicción con nuevos datos. [62]

2.8.4.3.- *K-nearest neighbors*

El algoritmo *k-nearest neighbors* (KNN) es considerado una metodología simple de *ML* supervisado que puede ser usado para solucionar problemas tanto de regresión como de clasificación. Este algoritmo permite reconocer patrones de forma rápida y detectar anomalías. De acuerdo con varios autores como Ligdi Gonzales (2017), este método es considerado un algoritmo no paramétrico basado en instancias, lo que significa que, el algoritmo no hace suposiciones explícitas sobre un dato, en lugar de esto, el algoritmo opta por memorizar las instancias de formación y usa estas como la base de su “conocimiento” para la fase de predicción. Este tipo de predicción solo se aplica cuando el usuario lo solicita para una etiqueta o parámetro de entrada específico. [63]

El KNN permite clasificar o calcular un parámetro de acuerdo con la distancia existente entre este parámetro y los datos del entrenamiento. En la figura 24 se muestra un ejemplo de clasificación a través de esta metodología.

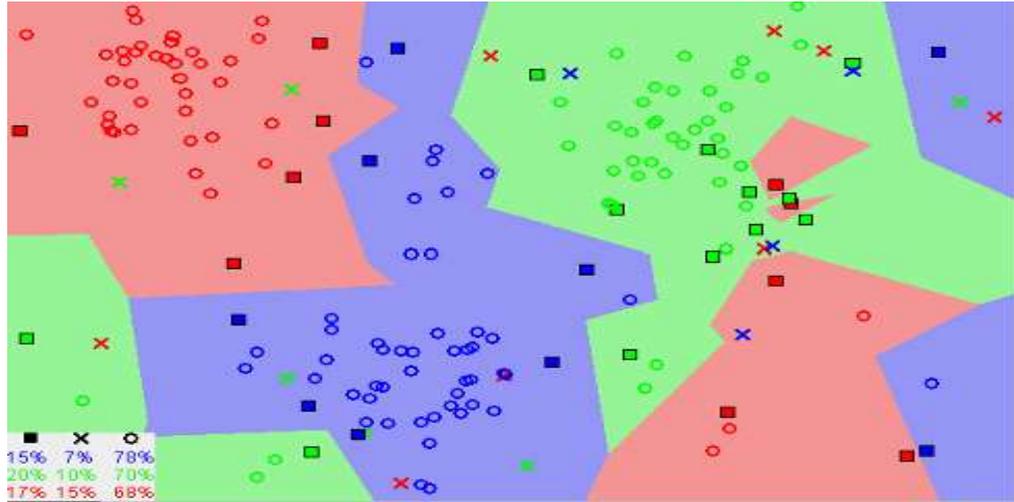


Figura 24: Ejemplo de K-Nearest Neighbors

Fuente: KNN <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d017>

Por lo general los algoritmos de KNN calculan la distancia con tres ecuaciones básicas, distancia Euclidiana, Manhattan y Minkowski, una vez calculadas las distancias, encuentra los puntos más cercanos y permite la clasificación o determinar el valor cuantitativo solicitado. [64]

2.8.4.4.- Árbol de decisiones

Los árboles de decisiones son uno de los algoritmos de *ML* más populares. Es un método analítico, que, a través de una representación esquemática, define las alternativas disponibles para la toma de decisiones. Su nombre se deriva de la apariencia del modelo parecido a un árbol. Este algoritmo de *ML* permite medir las predicciones logradas y valorarlas con los resultados obtenidos. Al final se realiza una comparación y se obtiene el árbol óptimo utilizando funciones predefinidas como los índices de Gini y la ganancia de información. A través de estas funciones, el algoritmo divide la información en nodos para obtener el primer nodo raíz y desde éste empieza a descomponer hasta obtener un estudio profundo de los datos de entrada. [65]

2.8.4.5.- Redes neuronales artificiales (*deep learning*)

Es un modelo simplificado inspirado en el funcionamiento de las neuronas en el cerebro cuando procesan la información. En el cerebro, las neuronas biológicas, reciben la información mediante las dendritas, luego, estos impulsos nerviosos se procesan en el soma y finalmente se transmite hacia otras neuronas a través del Axón que es la parte encargada de transmitir dichos impulsos. En una red neuronal artificial (RNA), cada nodo o neurona, recibe datos como entrada, estos valores son procesados y producen una salida, que luego se convierten en la entrada de las siguientes neuronas interconectadas. En la figura 25, se observa la comparación entre una neurona biológica y una neurona artificial. [66] [67] [68] [69] [70] [71].

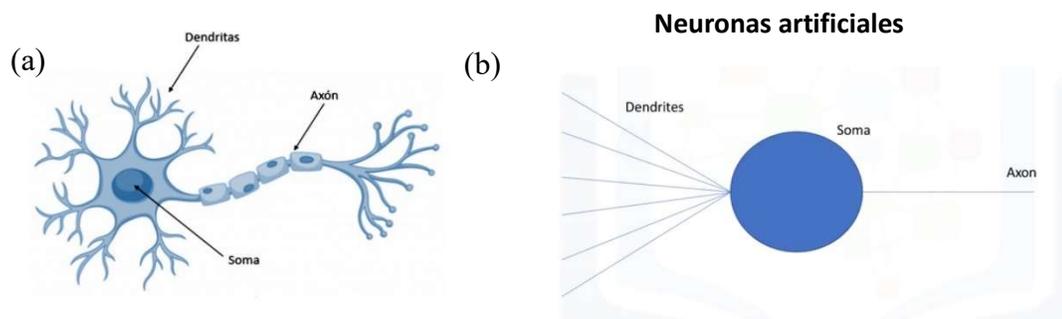


Figura 25: (a) Arquitectura de una neurona biológica (B) Arquitectura de una neurona artificial

Fuente: <https://www.xeridia.com/blog/redes-neuronales-artificiales-que-son-y-como-se-entrenan-parte-i>

- Arquitectura básica de una red neuronal

Las RNA consisten en una cantidad de nodos o neuronas agrupadas en capas, que se conectan entre sí. Una red neuronal debe contener al menos dos capas, una de entrada y una de salida. Normalmente las RNA poseen muchas capas, en estos casos, las capas que se encuentran entre la de entrada y la de salida se denominan capas

ocultas. A continuación, se explican cada uno de los elementos antes mencionados. [67]
[68]

- Capas: se le denomina así, a un grupo de neuronas artificiales cuyos valores de entrada vienen de una capa anterior y sus valores de salida pasan a ser los valores de entrada de la capa siguiente. En el caso de la capa de entrada (*input layer*), los valores de entrada de esas neuronas son los datos que se van a modelar, y en el caso de la capa de salida, las salidas de esas neuronas son los valores finales del resultado. [66] [67] [68]

- Capa de entrada (*Input Layer*): en esta capa es donde se ingresan los datos de entrada de la RNA, el número de neuronas en esta capa depende del número de parámetros de entrada que tenga la red. [60] [69]

- Capa de salida: en esta capa es donde se encuentran las neuronas que contienen los valores de salida de la red (*outputs*). El número de neuronas en esta capa es igual al número de valores que se desea estimar o predecir. [60] [69]

- Capas ocultas (*Hidden layers*): se les denomina así a todas las capas intermedias que existen entre la capa de entrada y la de salida. Su nombre proviene del hecho de que no se conocen ni los valores de entrada ni los valores de salida de cada una de ellas. El número de capas ocultas es variable en cada red. Una red puede contener desde cero capas ocultas (solo la capa de entrada y la de salida) o puede contener dos o más capas ocultas (a este último caso se le conoce como aprendizaje profundo o *Deep Learning*), y cada una de estas capas tiene un número de neuronas definido. [66] [69]

- Neurona Artificial: es una unidad de cálculo que se usa para la construcción de redes neuronales artificiales. A nivel esquemático, la neurona artificial se asemeja a una neurona biológica, en la que se pueden diferenciar 3 partes principales: los valores de entradas (que representarían a las dendritas), una función de activación (que representaría al Soma) y el valor o valores de salida (que representaría el Axón). El funcionamiento básico de una neurona es el siguiente: a cada valor de entrada, se le asigna aleatoriamente un valor denominado peso (*weight*), luego, se realiza una suma ponderada de todos los valores de entrada, es decir, se multiplican

cada uno de estos valores por el peso asignado y luego se suman, agregándole un valor constante llamado Bias. Una vez que se obtiene la suma, se aplica una función no lineal (denominada función de activación), y el resultado de esta función es el valor de salida que tiene la neurona. [66] [67] [68] [69] [72]

- Escalamiento: es otro de los parámetros fundamentales al momento de desarrollar este tipo de modelo, debido a que internamente se realizan muchas operaciones aritméticas sobre los datos de entrada. Cuando existe gran diferencia entre los valores de entrada, el tiempo de procesamiento suele ser mayor que cuando se usan valores mucho más pequeños.

Adicionalmente a la arquitectura, una RNA depende de diversas funciones internas.

- Funciones internas de una RNA

Entre las funciones internas de una RNA, tenemos la propagación directa, la propagación hacia atrás, la función de costo o pérdida, el descenso por gradiente, el entrenamiento y las curvas de aprendizaje. Todas estas son básicas para un correcto análisis de este tipo de modelo.

- Propagación directa (*Forward propagation*): es el proceso mediante el cual la información de entrada se va transmitiendo a través de las capas hasta llegar a la capa de salida. [73]

- Propagación hacia atrás (*Backward propagation*): es el proceso mediante el cual, una vez calculado el error entre los valores reales de salida (también conocido como *ground truth*) y los valores estimados/predichos, se usa este error para actualizar y recalcular los valores de los pesos y Bias de las neuronas en la red, con el fin de minimizar este error y hacer un mejor ajuste del modelo. [68] [74]

- Funciones de activación (*Activation function*): son funciones no lineales que se aplican al resultado de la suma ponderada de los valores de entrada de la

neurona. El resultado de esta transformación indica si la señal recibida por dicha neurona es relevante o no (en otras palabras, si la neurona es activada o no, de allí el nombre de este conjunto de funciones). Esto se hace con la finalidad de mejorar el desempeño de la red neuronal y que pueda realizar tareas más complejas. ^[69] ^[75]

○ Función de costo o de pérdida (*Cost function o loss function*): es un método o función que se usa para evaluar el desempeño de una red neuronal, o, dicho en otras palabras, es un método que sirve para cuantificar que tan bien se ajusta un modelo a un grupo de datos. En los problemas de optimización, lo que se busca es minimizar los valores resultantes de esta función. ^[76] ^[77] Entre las funciones de costo más usadas, se encuentran: *Mean Squared Error* o MSE y *Mean Absolute Error* (MAE).

El MSE se calcula al elevar al cuadrado todos los errores, y luego computar el valor promedio de los mismos.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad \text{Ec. 10}$$

Donde:

MSE: Mean Squared Error

n: Cantidad de datos

Y_i : Valores observados

\hat{Y}_i : Valores estimados

El MAE se define como el promedio del valor absoluto de la resta entre el valor real y el resultado de la predicción.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad \text{Ec. 11}$$

Donde:

MAE: Mean Absolute Error

n: Cantidad de datos

Y_i : Valores observados

\hat{Y}_i : Valores estimados

- Descenso por gradiente (*Gradient descent*): es un algoritmo iterativo de optimización, que se usó para poder hallar el mínimo de una función. En el caso de las redes neuronales, se usó para hallar el mínimo de la función costo. Este algoritmo calcula el negativo del gradiente de una función evaluada en un punto, luego, toma un segundo punto (que se encuentra en la dirección hallada anteriormente) y vuelve a evaluar la función en ese punto. Este proceso se realizó varias veces hasta que se logró hallar el mínimo, o un valor muy cercano a él. [76] [78]

- Entrenamiento: se le conoce como entrenamiento de una red neuronal, al proceso que se realiza con el fin de encontrar los valores de los pesos en las neuronas, que den como resultado unos valores de salida satisfactorios. El entrenamiento se debe realizar de manera supervisada, es decir, que de antemano se conozcan los resultados reales (también conocido como *groundtruth*) asociados a un grupo de valores de entrada. Una vez que se conozcan los valores reales de salida, se asignan valores aleatorios a los pesos y se realiza la propagación directa, luego, se calcula el error que hay entre los valores predichos (resultado de los pesos aleatorios) y los valores reales (*groundtruth*) usando la función costo, para luego optimizar estos resultados mediante el descenso por gradiente. El entrenamiento de una red neuronal se puede resumir en las siguientes etapas: [66] [68] [74]

- ✓ Inicializar los pesos y los Bias.
- ✓ Se calculan las salidas de la red mediante la propagación directa.
- ✓ Se calcula el error entre los valores predichos/estimados y los valores reales
- ✓ Se actualizan/recalculan los pesos y los Bias a través de la propagación hacia atrás
- ✓ Se repiten los últimos tres pasos hasta haber alcanzado un número máximo de iteraciones o hasta que el error haya alcanzado un límite preestablecido

- Curvas de Aprendizaje (*Learning Curves*): Son curvas que representan los valores que la función de pérdida (loss function) tiene a medida que se va entrenando una RNA. Normalmente en el eje de las abscisas se refleja la cantidad de

iteraciones (Epochs) establecidas para el entrenamiento, mientras que el eje de las ordenadas se presenta el valor de la función de pérdida que tiene el modelo para una determinada iteración. Existe una curva para los datos de entrenamiento y existe una curva para los datos de validación. Ambas se encuentran en la misma gráfica para poder estudiar el comportamiento de la RNA a medida que se realizan las iteraciones. En la figura 26, se observa el comportamiento ideal de una curva de aprendizaje, el cual se refiere a que ambas curvas tengan una tendencia decreciente y que, al momento de finalizar el entrenamiento, los valores de ambos errores sean muy parecidos y bajos.

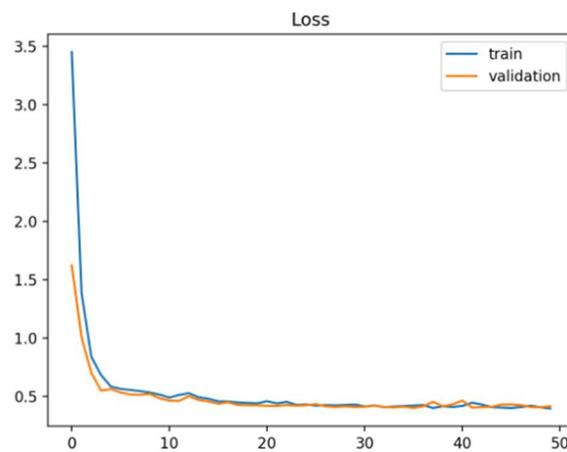


Figura 26: Ejemplo de buen ajuste, comportamiento ideal

Otro de los casos que se podría observar es que las iteraciones no sean suficientes para que las curvas de aprendizaje se estabilicen, sino que, por el contrario, al terminar las iteraciones, las curvas siguen decreciendo, como se observa en la figura 27a, en este caso se podría decir que la RNA está subentrenada, si se aumenta el número de epochs se podría llegar a un mejor resultado.

Usualmente, la curva de entrenamiento casi siempre tiende a decrecer, ya que al aumentar el número de iteraciones, el modelo se ajusta cada vez más a los datos de entrenamiento, pero, si la cantidad de iteraciones es muy alta, podría ocurrir que mientras la curva de entrenamiento siga disminuyendo, la curva de validación disminuya hasta cierto punto y luego la misma comience a crecer, como se observa en la figura 27b. Ésto ocurre porque el modelo se ajusta tanto a los datos de entrenamiento

que cuando se evalúa en datos distintos, el error aumenta. A esto se le conoce como sobreajuste (*overfitting*). El sobreajuste no es deseado, ya que la finalidad es obtener un modelo que permita describir el comportamiento de algún fenómeno, con el objetivo de poder usar esa información y predecir o estimar alguna variable, basado en posibles condiciones a futuro.

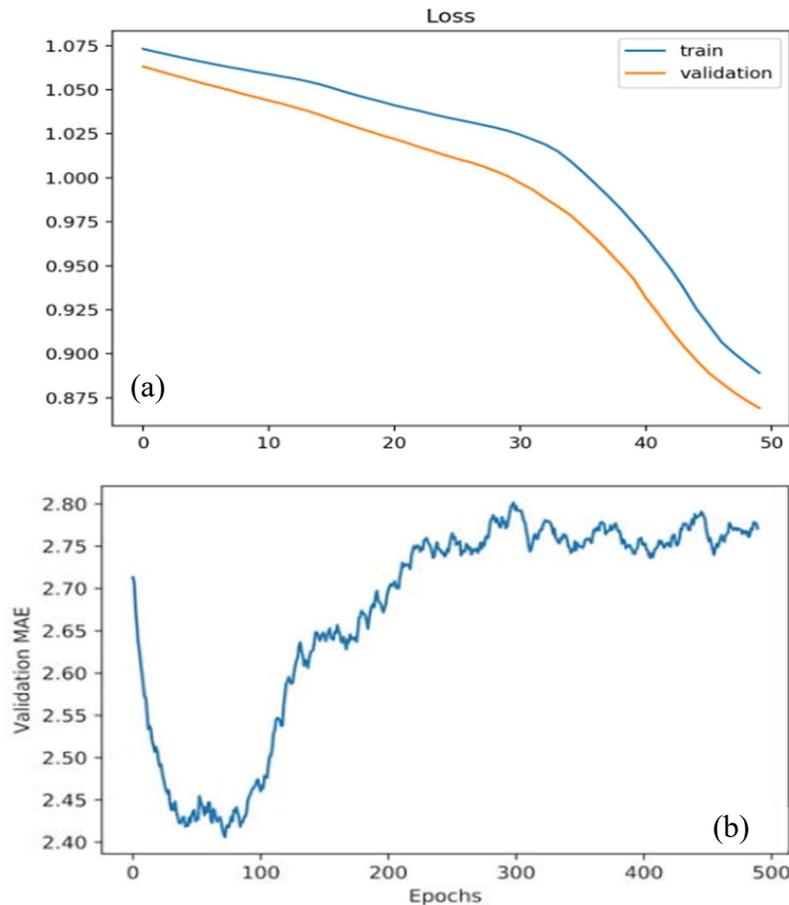


Figura 27 (a) Ejemplo de una curva de entrenamiento subentrenada (b) ejemplo de una curva con sobreajuste

Todos los conceptos y metodologías antes mencionadas son importantes, sin embargo, este TEG está basado en la metodología CRISP-DM, bajo el concepto de aprendizaje supervisado. Las disciplinas a utilizar son: la minería de datos la cual permitirá la validación de calidad de la base datos y una caracterización estadística adecuada, el aprendizaje automatizado, el cual permitirá la predicción de la producción

de petróleo, bajo dos modalidades, regresión lineal y polinomial, al igual que el análisis del comportamiento de las bombas, basado en metodologías de agrupamiento como lo son KNN o SVM, y la detección de fallas, a través de un árbol de decisiones. Por último, el aprendizaje profundo, basado en el uso de redes neuronales, los cuales permiten predecir dos parámetros al mismo tiempo, como lo son las tasas de petróleo y agua.

El último punto para tocar en este capítulo corresponde a la descripción del área de estudio, el cual es fundamental para entender el comportamiento del sistema pozo yacimiento, las lecturas de los sensores y posibles fallas presentes en el sistema.

2.9.- Descripción del área de estudio

La Faja Petrolífera del Orinoco (FPO) está situada en el extremo sur de la Cuenca Oriental, al norte del río Orinoco, comprende parte del Sur de los Estados Guárico, Anzoátegui, Monagas y Delta Amacuro, con un área de aproximadamente 55.000 Km² y las reservas probadas más grandes del mundo de petróleo pesado y extrapesado. El área de explotación actual comprende cuatro grandes bloques, Boyacá, Junín, Ayacucho y Carabobo.

El área de estudio de este TEG pertenece al bloque Junín, específicamente al área de PDVSA Petrocedeño como se muestra en la figura 28. La zona de estudio está ubicada en la localidad de San Diego de Cabrutica, municipio Monagas al sur del estado Anzoátegui con un área que abarca aproximadamente 399 Km² y se estiman

reservas probadas de aproximadamente 2,5 billones de barriles y una profundidad promedio de 2000 pies. [9]

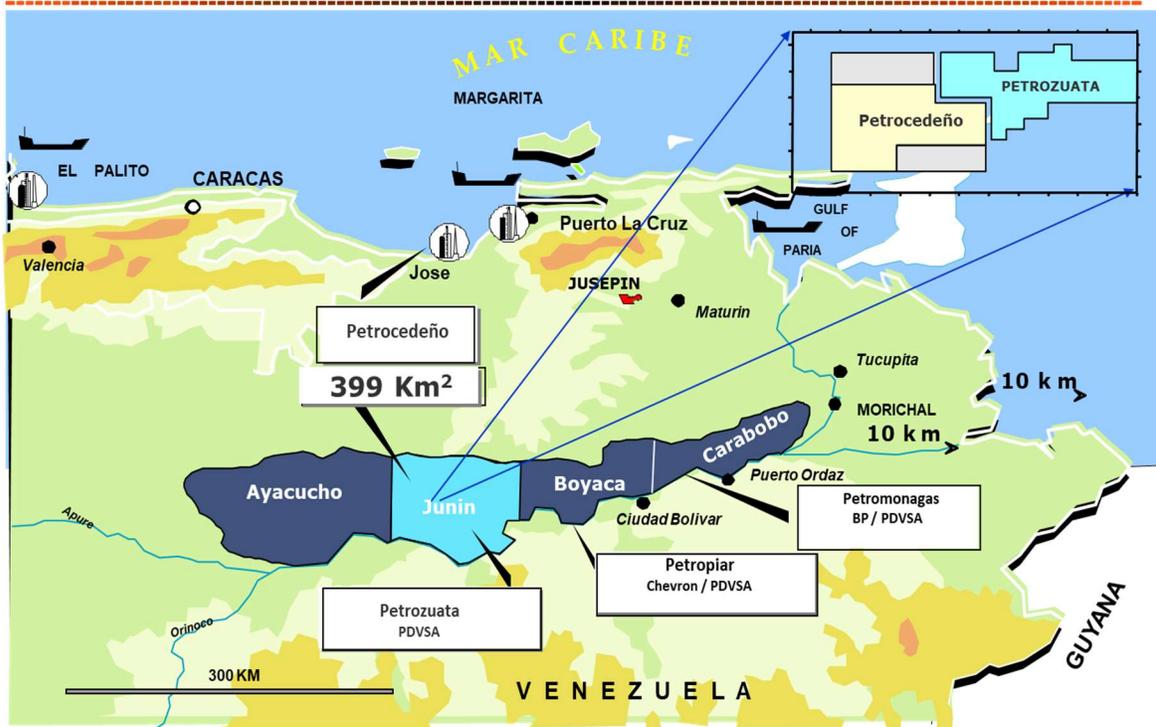


Figura 28: Ubicación del área de estudio en la FPO Bloque Junín.
Fuente: Statoil Petrocedeño. Procesos básicos de producción.(Octubre 2010.)

Los yacimientos de esta zona se caracterizan por ser yacimientos de arenas no consolidadas, provenientes de canales en formas de meandro formadas en su mayoría en la zona de deltaico y canales entrelazados en la zona fluvial como se muestra en la figura 29a. Esto genera que la producción de petróleo este acompañada de producción de arena.

Estos yacimientos producen bajo el mecanismo de expansión de gas en solución, sin embargo, en algunos pozos se aprecia un empuje hidráulico del acuífero presente, el cual debería constituir un mecanismo de producción adicional. Sin embargo, la presencia de este acuífero se hace más evidente en las arenas fluviales (ver figura 29a), en las cuales muchos pozos presentan evidencia de canalización de agua. La alta relación de movilidad genera un barrido no uniforme, la cual favorece el

adensamiento y causa una alta producción de agua en los pozos como se muestra en la figura 29b^[79], disminuyendo el factor recobro de esta zona.

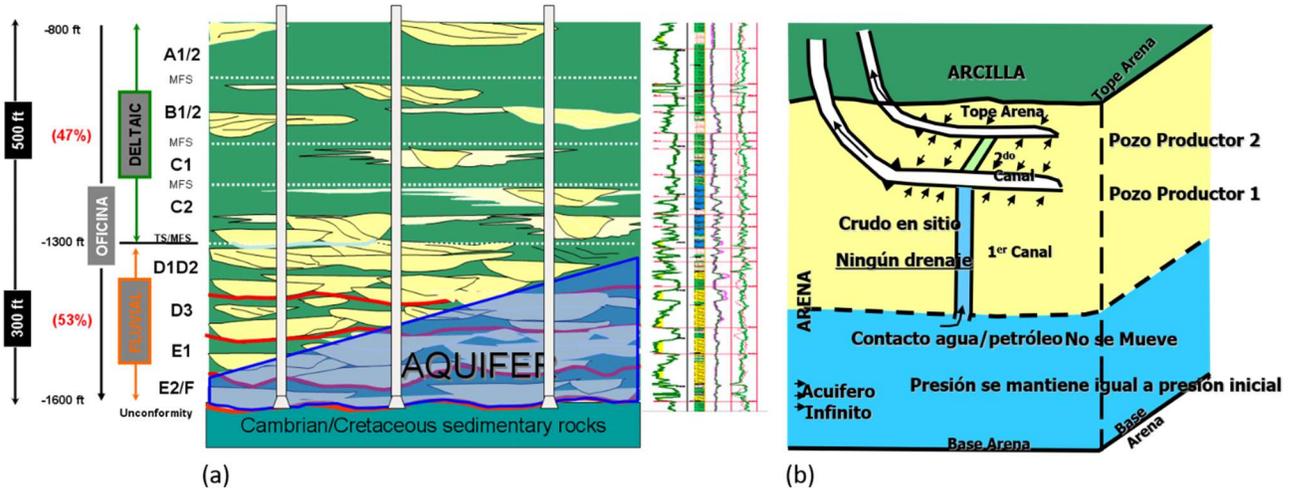


Figura 29: (a) Corte transversal del campo en estudio (b) Ejemplo de canalización del acuífero en los pozos horizontales del campo

Fuente: Statoil Petrocedeño. Petrocedeño upstream facilities.(Febrero 2013.)

El yacimiento en estudio se caracteriza por presentar petróleo extrapesado con una gravedad API promedio de 8,5° y viscosidades que van desde los 1000 hasta los 11000 cp aproximadamente. Estos yacimientos presentan una muy buena porosidad, con un promedio de 30% y una permeabilidad promedio de 10000 mD.

Para el año 2010, la empresa mixta Petrocedeño contaba con aproximadamente 40 macollas distribuidas en el campo y aproximadamente 1050 pozos en la zona. Cada macolla contiene un promedio de 29 pozos distribuidos entre los 9 yacimientos presentes. (A1/2, B1/2, C1, C2, D1/D2, D3, E1 y E2/F). Adicionalmente estas macollas se encuentran conectadas a una estación principal a través de tuberías las cuales se conocen como troncales. El campo cuenta con 4 troncales principales (T1, T2, T3 y T4), que, a su vez, se dividen en 2 como se muestra en la figura 30. Estas troncales son fundamentales para el transporte del petróleo desde la macolla hasta la estación principal y es necesario tenerlas siempre en cuenta para cualquier análisis o estudio en el campo, ya que, cualquier problema en este sistema ocasiona paradas e inconvenientes en las macollas conectadas a las mismas.

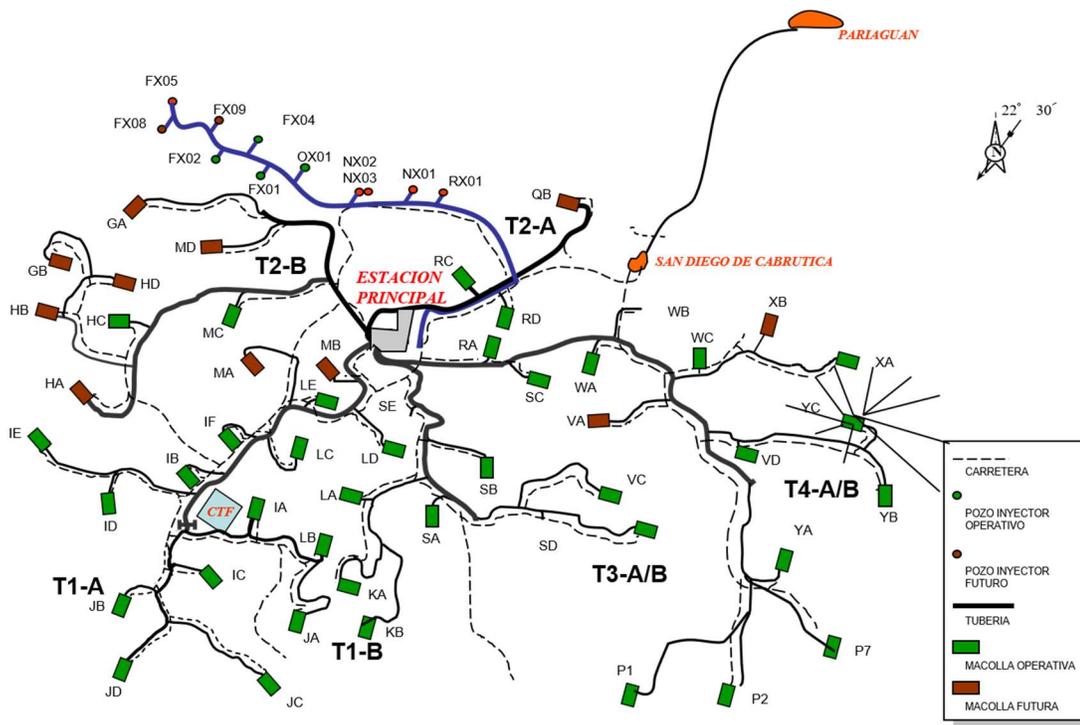


Figura 30: Distribución de las macollas y troncales en el campo en estudio.
 Fuente: Statoil Petrocedeno. Petrocedeno upstream facilities. (Febrero 2013.)

Para el año 2015, la compañía encargada del campo demostró a través de un análisis de producción de arena que las macollas tenían una alta concentración de arena, lo que afecta directamente al transporte de hidrocarburos, causando taponamiento a nivel de yacimiento y en los equipos de subsuelo y superficie. [80]

Las macollas están diseñadas para que los pozos estén distribuidos y no exista interferencia en sus áreas de drenaje. Por lo general los pozos son perforados paralelos con una distancia definida entre unos y otros, sin embargo, en la zona de estudio el arreglo de los pozos se presenta en forma de estrella, donde los pozos se encuentran juntos en el punto de partida y a medida que se alejan la distancia entre ellos aumenta.

[10] En la figura 31, se muestra la diferencia entre los arreglos de pozos antes mencionados.

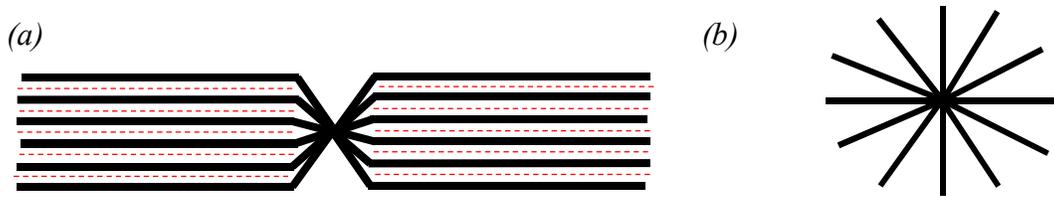


Figura 31: (a) Arreglo de una macolla de pozos horizontales en paralelo
(b) arreglo de una macolla tipo estrella

Cada macolla del campo está equipada con válvulas, medidores y bombas para poder transportar los fluidos. Las válvulas multipuertos le permiten al fluido desplazarse desde las líneas de flujo hacia el medidor o la bomba multifásica según sea el caso. Cuando los fluidos llegan a la bomba multifásica, éstos, son bombeados hasta la estación principal. En el área de estudio una macolla originalmente podía albergar hasta 24 pozos, a pesar de ello, en la actualidad existen macollas con hasta 50 pozos activos. En la figura 32 se muestra el diagrama de flujo de una macolla en el campo de estudio, donde se identifica la ubicación de todos los elementos antes mencionados. [9]

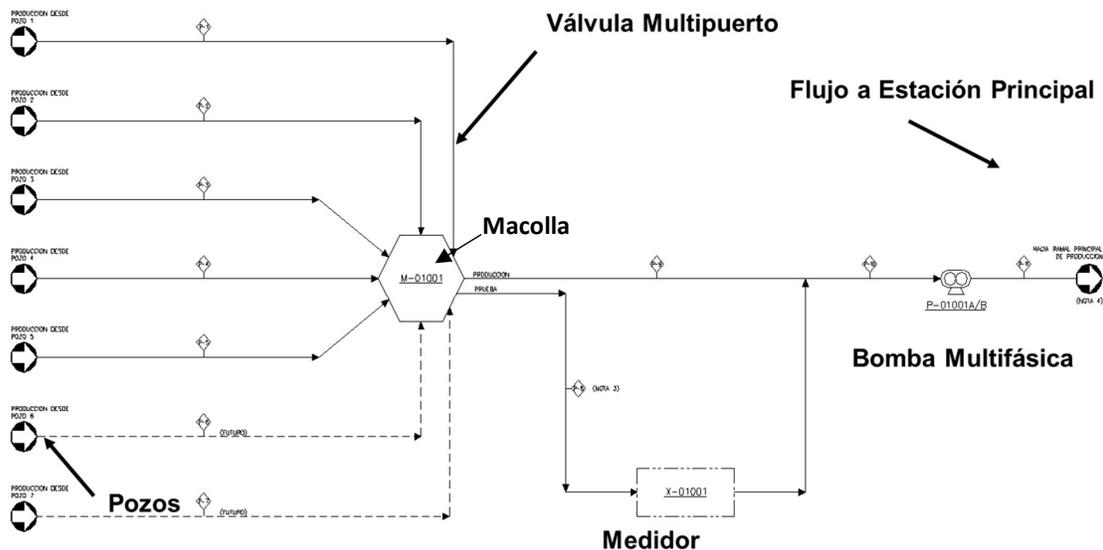


Figura 32: Diagrama de una Macolla
Fuente: Statoil Petrocedaño. Procesos básicos de producción. (Octubre 2010.)

La FPO en Venezuela se caracteriza por un diseño de producción de pozos horizontales. En este campo, desde el comienzo de la producción es necesario instalar mecanismos de levantamiento artificial aunado a la inyección de diluentes para poder disminuir la viscosidad del petróleo y mejorar el proceso de deshidratación. [25]

Cada pozo esta completado por elementos en superficie que controlan los elementos del fondo como lo son, las líneas de inyección de diluyente, el cabezal y los sensores de medición de flujo y presión. Por otro lado, están los elementos de subsuelo, como los revestidores y *liners*, sensores de fondo y las bombas de cavidades progresivas que son el sistema de levantamiento artificial. En la figura 33, se muestran los elementos del pozo. La mayoría de los pozos siguen este patrón de completación, con una longitud aproximada de 4500 pies, mientras que, la bomba de cavidades progresivas se encuentra en la sección tangencial del pozo.

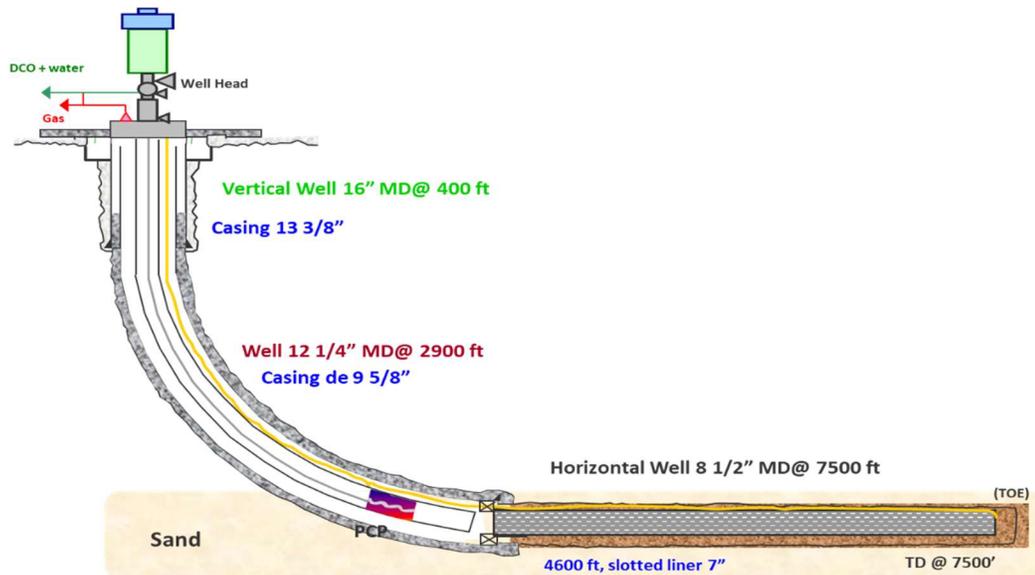


Figura 33: Completación típica de un pozo en el campo en estudio
Fuente: Statoil Petrocedeño. Procesos básicos de producción. (Octubre 2010.)

Una vez que el pozo empieza a producir, este es monitoreado desde el cuarto de control donde se realizan todas las verificaciones de la velocidad de la bomba, presión de cabezal e inyección de diluyente en el sistema. Todos estos datos son

almacenados para su posterior análisis por el ingeniero de yacimientos o producción encargado.

En el capítulo II, se han estudiado todos los conceptos teóricos tanto de yacimiento, mecanismos de producción, bombas de cavidades progresivas y la teoría básica de nuevas metodologías de *data science*, *machine learning* y *deep learning* los cuales serán fundamentales para el desarrollo del marco metodológico.

CAPITULO III

MARCO METODOLÓGICO

En este capítulo se describen todos los pasos que se realizaron para cumplir los objetivos de esta investigación.

La metodología de este trabajo está fundamentada en un tipo de investigación analítica y aplicada. Se considera una investigación analítica porque se fundamenta en la comparación y análisis de diferentes variables que pueden afectar la productividad de un conjunto de pozos, y aplicada porque busca desarrollar herramientas que faciliten la toma de decisiones en la empresa.

La primera fase de la metodología comprende la revisión bibliográfica y la recopilación de información que permitan seleccionar las metodologías y herramientas adecuadas para cumplir los objetivos previstos.

3.1.- Revisión bibliográfica y recopilación de información, y datos de interés

En esta primera fase de la investigación se establecieron las bases teóricas que sustentan todo el trabajo realizado, la cual abarca la revisión de *papers*, artículos de revistas, publicaciones de la Sociedad de Ingenieros de Petróleo (SPE), manuales y diferentes fuentes bibliográficas. Este procedimiento permitió conocer de manera más detallada el yacimiento, los sistemas de producción, el funcionamiento, aplicación y mantenimiento de las BCP, una descripción detallada del área de estudio, su estructura y facilidades de superficie y subsuelo, y adicionalmente toda la información asociada al uso de metodologías de *Data Science* en el mundo y en la industria petrolera.

La aplicación de este tipo de metodologías conlleva un gran aprendizaje tanto matemático, estadístico e informático, por lo cual para cumplir los objetivos propuestos fue necesario investigar conceptos básicos de estadística y desarrollar habilidades de programación con el lenguaje de Python. Un correcto análisis de datos requiere de una amplia manipulación de los mismos, donde Python fue una herramienta clave que

permitió hacer este análisis de forma eficiente y rápida por el gran volumen de datos. Por lo tanto, en esta primera fase fue necesario aprender a:

- Importar y exportar bases de datos de diferentes formatos. Este proceso permitió crear bases únicas en formatos predefinidos y de fácil acceso para el desarrollador.
- Manejar y organizar bases de datos. Este fue otro factor clave en el cumplimiento de los objetivos planteados, ya que se buscó reorganizar la información y descartar aquellos parámetros que no se consideraban importantes o útiles para la investigación. En este punto fue necesario crear y manipular estructuras de datos como, listas, diccionarios, matrices y otros que permitieron agrupar y clasificar los datos en base a distintas variables.
- Aplicar funciones matemáticas a una serie de datos y calcular parámetros de estadística básica, como: desviación estándar, percentiles, varianza, entre otros. La transformación de las bases de datos y cálculos estadísticos facilitó el proceso, disminuyendo el tiempo de cómputo y la creación de los algoritmos base.
- Aplicar funciones básicas de Python en base a cada librería importada. (Pandas, Numpy, Matplotlib, Plotly, SKLearn, Statsmodel, entre otras), Esto permitió crear y desarrollar algoritmos complejos de forma eficiente.
- La creación de gráficas interactivas con diferentes formatos de tiempo y fecha, facilitaron la visualización de los datos y ayudaron a concluir de manera eficiente.
- Conocer y entender cuál es el funcionamiento de cada una de las metodologías de *Machine Learning* y *Deep Learning*, al igual que sus aplicaciones. Este proceso permitió concluir cuales eran los métodos más efectivos para obtener los resultados deseados.
- Generar una interfaz gráfica de fácil acceso para los usuarios que facilitó el despliegue de todos los algoritmos creados.

Basados en la información recopilada, la metodología de *Data Science* que más se ajusta a este trabajo es la metodología CRISP-DM.

3.2.- Aplicación de la metodología CRISP-DM

En la figura 34 se muestra el esquema base de la metodología CRISP-DM aplicado a cualquier estudio de ciencias de datos. En esta segunda fase se adaptó cada una de las etapas de la metodología para cumplir los objetivos de la investigación.

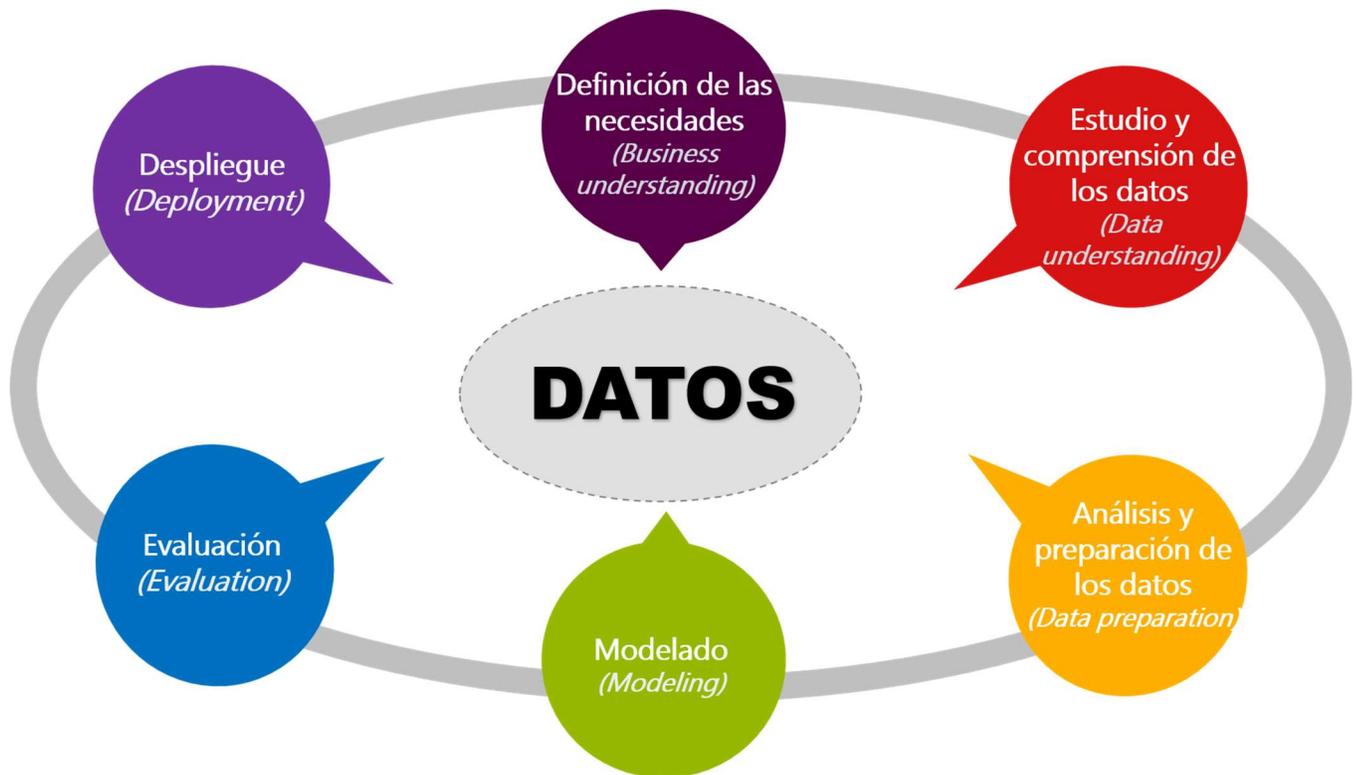


Figura 34: Etapas de la metodología CRISP-DM

3.2.1.- Definición de las necesidades o entendimiento del negocio (*Business Understanding*)

En la primera etapa de la metodología CRISP-DM, se analizaron los objetivos planteados en la investigación, con la finalidad de entender las necesidades y cuáles son los requerimientos para poder llevarlo a cabo.

En el análisis realizado se observó que la empresa mixta Petrocedeño ha presentado en los últimos años un aumento excesivo en la cantidad de servicios prestados a los pozos, consecuencia de una caída en la producción de petróleo e incrementos en sus costos operativos. Muchos de estos servicios, incluyen los cambios constantes y no planificados de bombas de cavidades progresivas, sin saber, si realmente la caída de producción es consecuencia de la bomba o del yacimiento. Esto ocasionó que existiese la necesidad de saber cuáles son los síntomas que presenta una bomba antes de dañarse o presentar cualquier falla y cómo se ven reflejados en el historial de producción. Para cumplir los objetivos de esta investigación, el estudio debe enfocarse en generar un modelo que describa la tendencia a lo largo del tiempo y definir cómo afectan los parámetros de producción y operatividad en el funcionamiento de la bomba, tomando en cuenta su marca o capacidad. Otra de las necesidades de la empresa, es saber si realmente la disminución de producción es causada por la bomba o si es un problema del yacimiento que amerita un trabajo de estimulación. El estudio de las bombas busca generar una herramienta para determinar si la bomba es eficiente bajo las condiciones operacionales del pozo o si, por el contrario, debe remplazarse o aplicarse un mantenimiento preventivo para extender su tiempo de vida.

Todo este planteamiento se englobó en la necesidad de optimizar el proceso de análisis de datos, para generar soluciones rápidas y óptimas a problemas operacionales o de yacimiento, con menores tiempos de análisis, menores costos y mayor eficiencia.

El análisis del problema aunado a la revisión bibliográfica y la recopilación de datos permitió avanzar a la segunda etapa de la metodología CRISP-DM, la cual, comprende el estudio y la comprensión de los datos.

3.2.2.- Estudio y comprensión de los datos (*Data Understanding*)

En esta etapa se seleccionó la muestra de estudio, sin embargo, el proceso de selección de la muestra fue complejo y ameritó el análisis de diversas variables como: ubicación y distribución del campo, cantidad de macollas, funcionamiento del sistema de producción, los procesos de transporte de fluidos desde cada macolla, arreglo de pozos en cada macolla y yacimiento, cantidad de pozos en la macolla, tipos de pozos presentes, completación del pozo, ubicación de sensores, datos recopilados en el historial de producción, tiempo de operatividad del pozo e información de servicios y bombas en cada pozo.

Este proceso de análisis y selección de muestra se dividió en tres fases: primero, el estudio de las bases de datos y material suministrado para la selección de la muestra, luego, el estudio, análisis y organización de las bases de datos y por último el análisis de la producción y operatividad de la muestra seleccionada basado en gráficos.

3.2.2.1.- Estudio de las bases de datos y material suministrado para la selección de la muestra

En la actualidad la empresa mixta Petrocedño no cuenta con un único formato para almacenar toda la información referente al campo, por lo tanto, la información no está almacenada de forma ordenada. Entre el material que se encontró se tienen archivos en Word, presentaciones de Powerpoint, archivos .PDF y existen muchas bases de datos en Access y vinculaciones con la aplicación *Oil Field Manager* de Schlumberger.

El análisis de los datos en las diferentes presentaciones permitió comprender el área de estudio, que, como se menciona en el capítulo II, consta de aproximadamente 40 macollas. Para la selección de la muestra se generó una tabla en formato Excel que recibe el nombre de “*DataMaster*”, con el cual se organizó y recopiló información general de cada macolla.

En la figura 35 se muestra la distribución del campo, donde cada uno de los círculos representa una macolla y las líneas que están alrededor representan los pozos de la misma. Estas macollas a su vez se encuentran conectadas a una troncal que distribuye los fluidos a la estación central. Con esta información se clasificó cada una de las macollas de acuerdo con su línea de distribución en cinco áreas, área principal central (PC), área principal este (PE), área principal oeste (PO), área norte (N) y área sur (S). Este dato fue fundamental para el análisis de datos, ya que, cualquier trabajo que se realice al oleoducto perteneciente a esta zona o a la planta principal generará una parada forzada en los pozos y macollas, lo cual debe verse reflejado en el historial de producción.

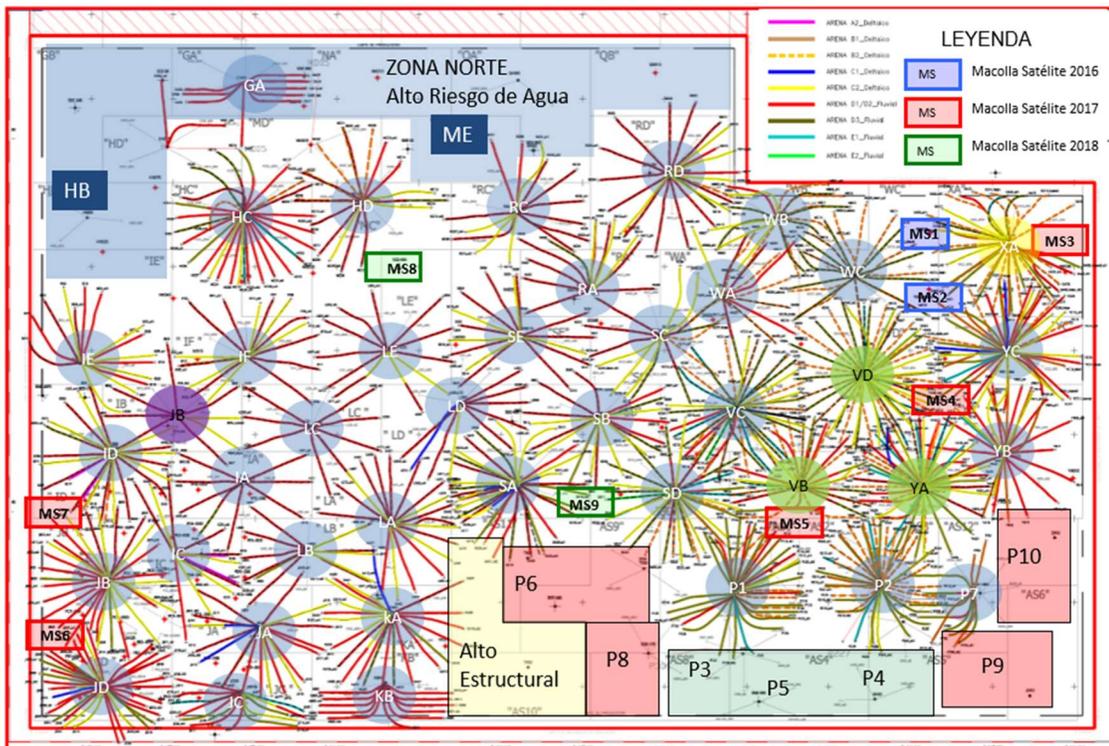


Figura 35: Mapa de la distribución de las macollas en el campo Petrocedeño
Fuente: Statoil Petrocedeño. Procesos básicos de producción.(Octubre 2010.)

El siguiente paso se basó en determinar cuántos pozos hay en cada macolla, por lo tanto, se comparó la cantidad de pozos existentes en cada una de las bases de datos (todas eran diferentes y presentaban diferentes periodos de tiempo y cantidad de pozos)

con los pozos representados en el mapa de la figura 35. Este procedimiento permitió no solo conocer el número total de pozos por macolla, sino también, determinar cuál base de datos es la más completa, algo fundamental para la investigación, ya que, la ciencia de datos se basa en la cantidad de datos que se tiene, por lo tanto, entre mayor es la cantidad de datos mayor será el aprendizaje de los algoritmos, obteniendo mejores resultados y un mayor alcance.

Al obtener la base de datos final, se procedió a analizar el historial de producción para cuantificar la cantidad de datos y días registrados por macolla y la fecha de inicio y final de producción. Todos estos datos se anexaron a la tabla antes mencionada, con los cuales se procedió a calcular: la cantidad de datos promedio que se tienen por pozo de cada macolla (este dato se obtiene dividiendo el total de datos de cada macolla entre su cantidad de pozos) y la cantidad total de días de producción de cada macolla (este dato se obtiene de la resta de la fecha final y la fecha inicial de producción de cada macolla). Al finalizar estos cálculos se procedió a totalizar, calculando la cantidad de datos total que se tiene por todas las macollas, el cual correspondió a 1.911.901 datos de producción. Por otro lado, se calculó la cantidad total de pozos de los cuales se tiene registro de producción, el cual correspondió a 829 pozos, se calculó la cantidad de datos promedio por pozo en cada macolla, el cual correspondió a 94.172 datos totales y por último se calculó la cantidad total de días de producción, el cual correspondió a 145.757 días. Con estos datos totalizados se procedió a encontrar el promedio de cada uno respecto al total de macollas, donde se obtuvo que el promedio de datos por macolla era de 50.313 datos, la cantidad de pozos promedio por macolla era de 22 pozos, la cantidad de datos promedio por pozo por macolla era de 2.478 datos y por último la cantidad promedio de días de producción por pozo por macolla, era de 3.836 días de producción.

Todos estos datos se representaron en la tabla de Excel, sin embargo, para definir la muestra final que cumpliera con los objetivos, aún, era necesario totalizar el número de servicios, cambios de bomba en cada macolla y el total de barriles producidos para jerarquizar según su rentabilidad económica. Para obtener la información de los servicios y cambios de bomba se revisó la base datos, al igual que

los archivos de Excel referentes a servicios. Este análisis permitió generar un solo archivo con todos los servicios a los pozos y se clasificaron por tipo de servicio y por pozo. Al finalizar de organizarlos, se totalizaron por macolla, para agregar los datos a la tabla de Excel “*DataMaster*”.

Por último, se contabilizó la cantidad de barriles producidos por cada macolla hasta la fecha para determinar la importancia y rentabilidad de la misma.

A los datos de servicio y producción se le realizó el mismo procedimiento antes expuesto para calcular la cantidad promedio de servicios y producción por macolla como se muestra en la tabla 3.

Con todos los datos organizados en Excel, se aplicó un formato condicional para determinar cuáles son las macollas que están por encima del promedio de todos los criterios, los cuales son: cantidad de datos disponibles, cantidad de pozos por macolla, cantidad de datos promedio por pozo, días de producción, cantidad de servicios realizados y producción total. Con estos criterios se logró clasificar las macollas en 3 grupos: las macollas de prioridad 1, son aquellas donde todos los criterios están por encima del promedio, las de prioridad 2, son aquellos que no cumplen con uno o dos criterios, y las de prioridad 3, son aquellas que no cumplen con más de dos criterios.

En la tabla 3 se muestra la tabla final con todos los datos calculados de las macollas con prioridad 1 y 2, donde en color verde se muestra todos los criterios que están por encima del promedio, dando como conclusión que las macollas con mayor número de criterios por encima del promedio, identificadas como prioridad 1, correspondieron a las macollas VD, VB y JB, siendo la de mayor importancia por su producción acumulada, cantidad de pozos, cantidad de datos promedio por pozo y mayor cantidad de servicios, la macolla VD, por lo tanto se selecciona como la muestra para el estudio a realizar.

Tabla 3: Análisis de datos para la selección de la muestra

MACOLLAS	Datos	Cant. Pozo	Cant. De datos Promedio por p.	Ubicació	Fecha de Inic	Fecha final	Días de producción	Servicios	Producción	Prioridad
VD	131303	47	2795	PE	11/12/2002	8/31/2013	3945	395	8.85E+07	1
VB	73798	28	2636	PC	12/1/2001	8/31/2013	4291	131	3.03E+07	1
JB	73313	28	2528	PO	12/1/2000	8/31/2013	4856	120	3.18E+07	1
XA	72366	31	2334	PE	1/1/2003	8/31/2013	3895	186	3.25E+07	2
YA	90043	38	2370	PE	4/1/2002	8/31/2013	4170	176	3.12E+07	2
VC	90056	42	2144	PC	12/1/2001	8/31/2013	4291	163	3.40E+07	2
SD	64699	30	2157	PC	12/1/2001	8/31/2013	4291	147	2.48E+07	2
YC	66044	41	1610.829268	PE	1/1/2003	8/31/2013	3895	137	3.11E+07	2
SA	67797	33	2054	PC	9/1/2001	8/31/2013	4382	130	1.78E+07	2
YB	58829	29	2028.586207	PE	6/1/2002	8/31/2013	4109	117	1.29E+07	2
WC	51922	17	3054	PE	1/1/2003	8/31/2013	3895	112	2.69E+07	2
HC	59423	29	2049	N	4/1/2002	8/31/2013	4170	101	2.84E+07	2
LA	77256	24	3219	PC	12/1/2001	8/31/2013	4291	94	2.94E+07	2
KA	54350	22	2470	PC	11/1/2001	8/31/2013	4321	85	2.20E+07	2
LB	55884	12	4657	PO	12/1/2000	8/31/2013	4656	81	2.49E+07	2
	Total de Datos	Total de Pozos	Cant. De Datos por Pozo				Tiempo total en días			
	1911901	829	94172				145757			
	Promedio de Datos por macolla	Promedio de Pozos	Promedio de Datos				Promedio de Tiempo			
	50313	22	2478				3835.71053			

Al seleccionar la muestra, el estudio, análisis y organización de las bases de datos se volvió más finito y concreto.

3.2.2.2.- Estudio, análisis y organización de las bases de datos

Para el estudio de las bases de datos fue necesario crear un archivo generalizado que incluyera toda la información de todos los pozos y macollas, ya que, este procedimiento facilitó la programación de los algoritmos principales.

La primera parte consistió en importar todos los pozos y asignarlos a cada macolla según corresponda, luego, se ubicó cada una de las macollas con sus coordenadas y se reportó en una tabla de información generalizada. Adicionalmente, se realizó la clasificación exacta de las macollas por troncal, ya que existen 4 troncales principales (troncal 1: T1, troncal 2: T2, troncal 3: T3, troncal 4: T4) con las cuales se separó la muestra inicial, pero existen a su vez 4 subdivisiones (troncal 1 A: T1A,

troncal 1 B: T1B, troncal 2 A: T2A y troncal 2 B: T2B) de las cuales se debe estar al tanto para un buen análisis de datos.

Al finalizar de organizar la información de las macollas, se procedió a realizar un estudio detallado de los pozos, con la finalidad de agregar al archivo, información acerca del tipo de pozo (si es productor de petróleo, un pozo exploratorio o de inyección), el ambiente sedimentario y el yacimiento o arena al que pertenecen. El ambiente sedimentario es de mucha importancia en este estudio, ya que, está relacionado con el tipo de mecanismo de empuje presente y será de gran impacto en el análisis de datos que se hará posteriormente, y por último se agregó la fecha de inicio de producción de cada pozo. En la tabla 4, se observa la estructura de la tabla de datos creada, donde se muestran todos los criterios de clasificación antes mencionados.

Tabla 4: Tabla generalizada con información principal de cada uno de los pozos

CLUSTER	WELLNAM	LATITUD	LONGITUD	TRUNKLIN	WELLYPE	RESERVOIR	ASSET	START_DATE
VD	VD01	8.358637632	-64.87186506	T4	OIL	B2	DELTAIC	11/12/2002
VD	VD02	8.358637632	-64.87186506	T4	OIL	D3	FLUVIAL	11/12/2002
VD	VD03	8.358637632	-64.87186506	T4	OIL	B2	DELTAIC	11/12/2002
VD	VD04	8.358637632	-64.87186506	T4	OIL	D3	FLUVIAL	11/12/2002
VD	VD05	8.358637632	-64.87186506	T4	OIL	B2	DELTAIC	11/12/2002
VD	VD06	8.358637632	-64.87186506	T4	OIL	D3	FLUVIAL	11/12/2002
VD	VD07	8.358637632	-64.87186506	T4	OIL	B2	DELTAIC	11/12/2002
VD	VD08	8.358637632	-64.87186506	T4	OIL	D3	FLUVIAL	11/12/2002
VD	VD09	8.358637632	-64.87186506	T4	OIL	B2	DELTAIC	11/12/2002
VD	VD10	8.358637632	-64.87186506	T4	OIL	C2	DELTAIC	11/12/2002
VD	VD11	8.358637632	-64.87186506	T4	OIL	D3	FLUVIAL	11/12/2002
VD	VD12	8.358637632	-64.87186506	T4	OIL	B2	DELTAIC	11/12/2002

Al finalizar la tabla generalizada se procedió a estudiar y analizar el historial de producción. Esta base de datos está conformada por 19 columnas, las cuales se detallan a continuación en la tabla 5. La mayoría de estos conceptos se explicaron en el capítulo II, sin embargo, es importante tomar en cuenta que se debe analizar cómo se realiza la adquisición de los datos de cada uno de los parámetros aquí mencionados.

Tabla 5: Descripción de los parámetros en historial de producción

Parámetros	Descripción	Unidades
<i>WELLNAME</i>	Nombre del pozo	N/A
<i>DATE</i>	Fecha de producción	aaaa/mm/dd
<i>DURATION</i>	Duración o tiempo en la que estuvo activo el pozo en el día	hrs.
<i>WHP</i>	Presión de cabezal	psi
<i>WHT</i>	Temperatura de cabezal	°F
<i>RPM</i>	Revoluciones por minuto de la bomba	rpm
<i>TORQUE</i>	Torque en las cabillas de la bomba	
<i>INTAKE PRESS</i>	Presión en la entrada de la bomba	psi
<i>DISCHARGE PRES</i>	Presión de descarga (a la salida de la bomba)	psi
<i>DISCHARGE TEMP</i>	Temperatura de descarga (a la salida de la bomba)	°F
<i>DILUENT VOL</i>	Volumen de diluyente inyectado	bpd/día
<i>GAS RATE</i>	Tasa de gas (producción de gas en el pozo)	ft3/día
<i>WATER RATE</i>	Tasa de agua (producción de agua en el pozo)	bpd
<i>PI</i>	Índice de productividad	
<i>PUMP EFFICIENC</i>	Eficiencia de la bomba	N/A
<i>ALLOCPRUEBA DATE</i>		
<i>STATUS</i>	Estatus del pozo	
<i>EHO RATE</i>	Tasa de petróleo (producción de petróleo)	bpd/día
<i>CASING PRESS</i>	Presión en el revestidor	psi

En la tabla anterior, se mostró una breve descripción de los parámetros en el historial de producción. A continuación, se explica cuál es el proceso de medición en el campo y la importancia de cada uno de los parámetros:

- *WELLNAME* (Nombre del pozo): Es un parámetro de clasificación, el cual es fundamental ya que permite identificar a que pozo pertenecen los datos adquiridos, este dato es asignado por la gerencia del campo y es único para cada pozo.
- *DATE* (Fecha): Este parámetro permite registrar los datos con una cronología, y se establece en formato (dd/mm/aaaa) calendario desde el día inicial de producción, el cual permite registrar los datos diariamente.

- *DURATION* (Duración): La duración es un parámetro medido en superficie el cual contabiliza la cantidad de tiempo diario en el cual el pozo se encuentra en producción. Este parámetro es medido en horas, por lo tanto, su rango está entre 0 y 24 horas.
- *WHP* (Presión de cabezal): su definición se explicó en el capítulo II. Este parámetro es medido a nivel de superficie a través de sensores de presión, estos son instalados en el cabezal del pozo y se controlan por medio de una bomba multifásica la cual permite mantener esta presión baja. Es importante tomar en cuenta que la presión de cabezal interactúa con la presión de fondo por medio del nivel de fluido, el cual es fundamental para mantener el pozo operativo, ya que la bomba debe permanecer sumergida en líquido. Adicionalmente la unidad de medida con la cual registra el sensor está en unidades de psi, por lo tanto, la presión mínima aceptada para este parámetro es de aproximadamente 14.7 psia, la cual corresponde a la presión atmosférica.
- *WHT* (Temperatura de cabezal): su definición se explicó en el capítulo II. Este parámetro se obtiene a través de un medidor de temperatura que se encuentra instalado en el cabezal del pozo. Este medidor se encarga de registrar la temperatura con la cual llegan los fluidos extraídos a la superficie. La unidad de medida de este sistema es el grado Fahrenheit (°F), por lo tanto, la temperatura mínima que puede ser aceptada para este parámetro es de aproximadamente 32 °F, la cual corresponde a la temperatura de congelamiento
- RPM (Revoluciones por minuto de la bomba): este parámetro es controlado por el operador del pozo desde el cuarto de control. Su comportamiento se estudia a través de los archivos de “*pump monitor*”, en el cual se registra el comportamiento de la bomba en el día, y al finalizar el mismo se reporta en la base de datos las RPM promedio con las cuales trabajó la bomba. El valor mínimo que debe reportarse es de 0 RPM.
- TORQUE: se define en el capítulo II. Es un parámetro medido a nivel de la bomba BCP, el cual es registrado a lo largo del día en el software de la bomba,

y al finalizar el día se reporta en la base de datos el torque promedio. Al igual que las RPM, el valor mínimo medido debe ser 0.

- *INTAKE_PRESS* (Presión de entrada de la bomba): esta presión se mide a través de sensores de fondo en la BCP, los cuales miden la presión con la cual entran los fluidos a la bomba. No se conoce con exactitud el valor de la presión mínima, ya que es un parámetro que disminuye respecto al tiempo de producción, sin embargo, debido a la profundidad en la que se encuentra el sensor, esta presión debe ser mayor a la presión atmosférica.
- *DISCHARGE_PRESS* y *CASING PRESS* (Presión de descarga de la bomba y presión de revestidor respectivamente): estas presiones se miden a través de sensores o *sonolog*, los cuales se encuentran en la parte superior de la BCP como se observa en la figura 36. Éste actúa como un doble sensor que mide tanto la presión de fondo del lado del revestidor, como la presión de salida de la bomba. Este sensor se encuentra sumergido en líquido y la presión medida del lado del revestidor se encuentra relacionada con la presión de cabezal, ya que, un pozo solo puede operar con seguridad cuando el nivel de líquido está a 100 pies por encima del sensor de fondo cuando este funciona, o 200 pies si el sensor no funciona, de lo contrario, hay riesgo de que la bomba se apague y deje de producir. Por otro lado, el sensor mide la presión a la salida de la bomba del lado del *tubing*, considerando esto, la presión de descarga, la cual debería ser mayor a la presión de entrada antes explicada, ya que mide la presión ejercida por toda la columna de líquido fluyendo que es transportada hasta la superficie.

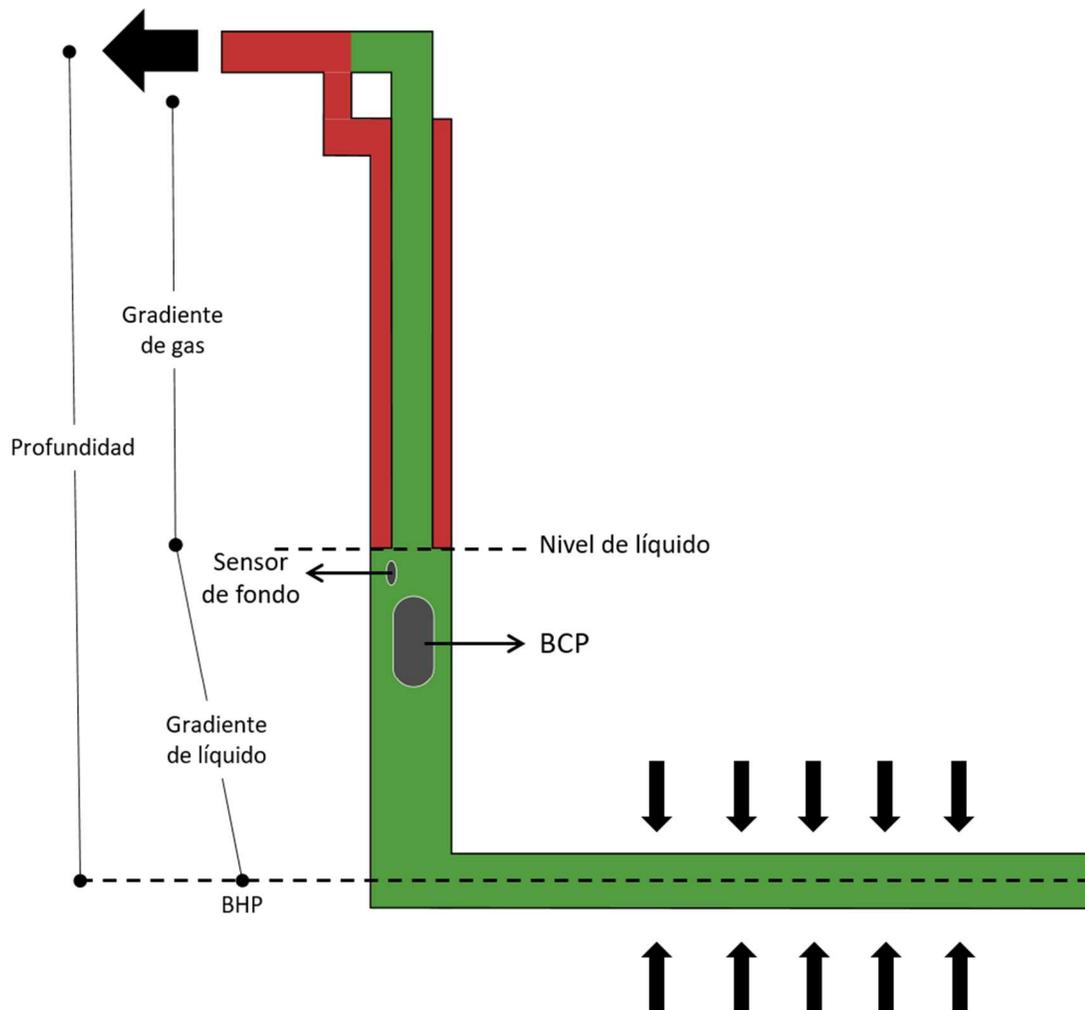


Figura 36: Ubicación del sensor de fondo encargado de medir la presión de salida de la bomba y la presión de revestidor

Fuente: Statoil Petrocedeño. Procesos básicos de producción. (Octubre 2010.)

- *DISCHARGE_TEMP* (Temperatura de descarga): es un parámetro que se mide a la misma profundidad a la que se encuentra el sensor de fondo y registra la temperatura con la cual pasan los hidrocarburos en este punto. Esta temperatura debe ser mucho mayor a la temperatura ambiental y es registrada en °F.
- *DILUENT_VOL* (Volumen de diluyente): este parámetro es medido desde superficie, con un medidor de flujo a través del múltiple de diluyente y controla la cantidad de volumen de diluyente inyectado al pozo con la finalidad de reducir la viscosidad del crudo.

- **PI (Índice de productividad):** es un parámetro calculado basado en la teoría explicada en el capítulo II. Este parámetro involucra la tasa de petróleo, la presión de fondo y la presión promedio en el área de drenaje, esta última tiene gran incertidumbre ya que no es medida directamente, sino que es estimada. Por esta razón, el índice de productividad se calcula basado en la presión inicial del pozo, lo que no permite calcular el potencial real del pozo con el tiempo, y se procede a construir un modelo *Pdrain*. El modelo *Pdrain* consiste en realizar una declinación lineal de la presión del área de drenaje respecto a la producción acumulada, y ésta es calibrada con las estimaciones de la misma, basados en el comportamiento de restauración de presión. El IP es recalculado después de cada prueba usando la nueva presión estimada, siendo un valor calculado que no tiene un rango específico y depende de la presión de fondo y la presión estática, sin embargo, éste nunca debe ser negativo.
- ***PUMP_EFFICIENC* (eficiencia de la bomba):** es otro parámetro calculado, el cual considera el fluido total levantado por la bomba (parámetro medido) en las condiciones que opera respecto al desplazamiento ideal por cada RPM. Es importante tomar en cuenta que la eficiencia de la bomba no solo disminuye por un daño ocasionado en la misma, sino, por la presencia de gas que causa que el volumen ocupado por el gas aumente y la eficiencia líquida se reduzca. De acuerdo con los estatutos de la empresa, si la eficiencia permanece constante en el tiempo y existe un cambio brusco en este parámetro se justifica una intervención en el pozo. Por ser una relación este valor se encuentra en el rango entre 0 y 1.
- ***ALLOCTEST_DATE*:** indica la última fecha en la cual se realizó una prueba o medición de la tasa de los fluidos en el medidor multifásico.
- ***STATUS* (estatus):** este parámetro indica si el pozo está activo o inactivo en esa fecha
- ***GAS_RATE* (tasa de gas):** este parámetro indica la cantidad de gas producido a lo largo del día. La unidad con la cual se reporta este parámetro es en pies

cúbicos normales por día (PCN/d) o también conocido como *standard cubic feet per day* (SCF/d).

- *EHO_RATE* (tasa de petróleo): indica el volumen de petróleo producido a lo largo del día y se reporta en barriles (bbl/día).
- *WATER_RATE* (tasa de agua): indica el volumen de agua que produce cada pozo a lo largo del día y se reporta en barriles (bbl/día).

Los valores de tasa registrados en las bases de datos nunca deben ser negativos. Antes de continuar con el análisis de las bases de datos, es necesario explicar con detenimiento, como es el procedimiento que se realiza en campo para medir las tasas de los fluidos producidos, ya que, cualquier error en la medición de estos volúmenes puede generar mayor incertidumbre en los resultados que se pueden obtener en este TEG.

Los caudales se estiman a través de los medidores multifásicos que en principio fueron los ROXAR y luego migraron a otra empresa llamada AGAR (fijos y móviles), contratados o propios, mientras el corte de agua (%AyS) se mide con muestras a pozos. Para el año 2014, de acuerdo con datos de la empresa, de un total de 5444 pruebas medidas, tan solo 2926 pudieron validarse, lo que equivale al 54%. Es necesario asegurarse que las pruebas cargadas en las bases de datos se realicen correctamente y además correspondan al comportamiento esperado del pozo, si éstas no son validadas aumenta la incertidumbre en los caudales reportados. Por otro lado, en caso de que no pueda realizarse una prueba real en más de 1 mes, se realiza una prueba estimada de producción o también conocida como “*Dummy test*”. Ésta considera las RPM a la cual se realiza la prueba, la presión para el cálculo del volumen de gas, asume las densidades de los fluidos en la última prueba y la tasa de diluyente inyectada por lo general en el pozo. Con todos estos datos se estima la tasa de petróleo, el corte de agua y la RGP, de igual forma como se esperaría para validar una prueba. Este último procedimiento genera un aumento en la incertidumbre de las tasas reportadas, algo perjudicial para un correcto análisis de datos.

Una vez que se estudiaron los procedimientos de medición y se analizó el historial de producción con detalle, se procedió a organizar y analizar los servicios. Este procedimiento se estructuró en tres pasos:

- Revisión y clasificación de los datos referente a los servicios: en la base de datos principal se ubica la tabla de los servicios de todos los pozos. La tabla de servicios está estructurada en 10 columnas de acuerdo con lo que se muestra en la tabla 6. Estos datos se exportaron a Excel para que el manejo y organización fuera más fácil.

Tabla 6: Descripción de los parámetros en la base de datos de los servicios

	Descripción
<i>WELLNAME</i>	Nombre del pozo
<i>LEGAL_WELLNAME</i>	Nombre legal de pozo con el que está registrado ante el ministerio
<i>Event_Type</i>	Tipo de servicio realizado
<i>Event_Objective</i>	Objetivo del servicio
<i>Start_Date_Of_Event</i>	Fecha de inicio del servicio
<i>End_Date_Of_Event</i>	Fecha final del servicio
<i>Pump_Model</i>	Modelo de la bomba
<i>TMD</i>	Profundidad medida a la cual se realizó el servicio
<i>Sensor_TMD</i>	Profundidad medida a la que se encuentra el sensor
<i>Sensor_TVD</i>	Profundidad vertical a la cual se encuentra el sensor

El primer paso fue filtrar los datos, con la finalidad de obtener sólo los servicios de la muestra seleccionada. Una vez terminado este procedimiento, basados en la descripción de cada uno de los servicios y en la asesoría de la tutora académica, fue necesario organizar y normalizar cada uno de ellos. Por esta razón, se generó un nuevo campo que permitiera agrupar los servicios en seis categorías: *Completion*: incluye todos los servicios de completación realizados a los pozos, *well optimization*: optimización de pozo, incluye los servicios relacionados con averías en el sistema del pozo o con procesos de optimización de bomba, *pump change*: incluye todos los cambios de bomba realizados, *stimulation*: incluye las estimulaciones realizadas al pozo, *well suspension*: suspensión del pozo, *other well services*: incluye cualquier otro servicio

realizado al pozo, como por ejemplo la pesca de herramientas, cambios de cabezal o cambios de sensores de fondo, adquisición de pruebas PVT, entre otras.

Terminado el proceso de categorización, se procedió al segundo paso.

- Estudio inicial de las bombas utilizadas en el campo: en esta fase se procedió a revisar todos los servicios referentes a las bombas, con la finalidad de obtener un registro de todas las marcas y modelos usados. Con esta información se realizó una investigación vía web, con la finalidad de obtener todos los catálogos correspondientes a las empresas fabricantes de las bombas y estudiar las especificaciones técnicas. Con estos catálogos se procedió a generar una tabla de datos por cada una de las marcas, estas tablas almacenan la información básica extraída de los mismos como: marca, modelo y tasas de flujo por RPM. Una vez estructurados estos datos se calculó el desplazamiento por cada una de las RPM para poder utilizarse en cálculos futuros. (Anexo 4)

Por último, se procedió a realizar el tercer paso.

- Compilación de los datos de las bombas y estandarización de modelos y marcas: en el momento de revisar exhaustivamente la base de datos, se observó que la mayoría de los modelos y marcas registrados no tenían un formato definido, algunos tenían nombres incompletos o faltaba el dato, lo que genera un aumento en la incertidumbre al momento de analizar los datos, sin embargo, se procedió a modificar todos los registros existentes con la finalidad de normalizar y colocar todos los nombres y marcas con el mismo formato. Este paso fue necesario, ya que, los datos deben tener un único formato para que puedan ser leídos de forma autónoma por el software.

Al entender y definir las bases de datos a utilizar, se procedió a realizar un análisis de la producción y operatividad del campo, basados en gráficos, con la finalidad de estudiar anomalías o tendencias en los datos.

3.2.2.3.- Análisis de la producción y operatividad del campo basado en gráficos

- Análisis de servicios

La primera parte de esta etapa se fundamentó en analizar cuáles fueron los servicios realizados en cada uno de los pozos de la muestra. Cada tipo de servicio genera un impacto diferente en el historial de producción, por lo tanto, se programó una herramienta que permitió clasificar la cantidad de servicios de acuerdo con la categoría creada en cada uno de los pozos. En la figura 37, se muestra la tabla de resultados obtenidos del algoritmo diseñado, donde se observa que la mayoría de los servicios corresponden a cambios de bomba, lo cual, corresponde con el problema analizado en la etapa 1.

	Completion	Well Optimization	Pump change	Stimulation	Other well services	Well suspention	CLUSTER	RESERVOIR
VD18	1	0	17	0	1	1	VD	D3
VD24	1	0	13	0	0	0	VD	D3
VD20	1	0	12	0	0	0	VD	D3
VD13	1	2	11	0	1	0	VD	D3
VD02	1	1	9	0	0	0	VD	D3
VD04	1	1	8	0	0	0	VD	D3
VD06	1	0	8	0	0	0	VD	D3
VD08	1	2	8	0	0	0	VD	D3
VD11	1	0	8	0	1	0	VD	D3
VD29	2	1	8	0	0	0	VD	D3
VD05	1	0	7	1	0	0	VD	B2
VD21	1	0	7	0	2	1	VD	B2
VD34	1	0	6	1	0	0	VD	C2
VD28	1	1	6	0	0	0	VD	E1
VD03	1	0	6	0	0	0	VD	B2
VD12	3	0	6	0	0	0	VD	B2
VD23	1	0	5	0	0	0	VD	D3
VD09	1	1	5	0	1	0	VD	B2
VD14	1	1	5	0	2	0	VD	B2
VD17	2	0	5	0	1	1	VD	B2

Figura 37: Tabla de Clasificación de Servicios por pozo de la macolla en estudio

Basado en estos resultados se debió separar los pozos en dos grupos, el primer grupo aquellos pozos con cambios de bomba y otros servicios, pero, que no han pasado por un proceso de estimulación, y en el segundo grupo, aquellos pozos que si han pasado por un proceso de estimulación. Esta separación permitió estudiar con detenimiento, como afecta el cambio de bombas en la producción sin ninguna otra perturbación en el sistema.

Para poder concretar el análisis de la producción se debió calcular nuevas variables y unificar las bases de datos, con la finalidad de crear una sola base de datos que permitiese a los algoritmos diseñados acceder a toda la información.

- Cálculo de variables adicionales necesarias para el análisis

Para el análisis de la producción fue necesario calcular parámetros adicionales en la base de datos del historial de producción, relacionados con las tasas, como: la tasa de líquido total producida diariamente (este parámetro se calcula con la ecuación 2 explicada en el capítulo II), los resultados obtenidos se añadieron al historial de producción como una nueva columna, la cual recibe el nombre de “*LIQUID_RATE*”. Otro parámetro importante es la relación agua petróleo (la cual se calcula con la ecuación 4 explicada en el capítulo II), los resultados obtenidos de este cálculo se agregaron en una columna que recibe el nombre de “WOR” (según sus siglas en inglés *Water Oil Ratio*). Por último, se agrega la relación gas petróleo (calculado con la ecuación 5) y la columna resultante recibe el nombre de “GOR” (según sus siglas en inglés “*Gas Oil Ratio*”).

Por otro lado, se calcularon parámetros derivados de la fecha, que facilitaron el análisis de los datos de acuerdo con un período seleccionado. Entre las columnas agregadas están: año, mes y día por separado, al igual que el tiempo de producción “t”.

Al terminar el procedimiento de cálculo de nuevas variables, se procedió a unir toda la información.

- Unión de bases de datos y análisis del historial de producción y comportamiento de bombas

El historial de producción, el archivo máster generado y el archivo de servicios se deben unir para generar una base de datos completa. La unión entre el historial de producción y los servicios se realizó tomando en cuenta el nombre del pozo y la fecha. Para lograr esto, se debió programar una función que filtrara ambas bases de datos por pozo y encontrara las fechas en la cual se realizan los servicios, el programa tiene la capacidad de agregar a la base de datos principal, que en nuestro caso es el historial de producción, toda la información de los servicios de manera ordenada. El resultado de esta nueva base generó un archivo que comprendía todo el historial de producción y los servicios en cada pozo.

El archivo master sólo contiene variables de clasificación, por lo tanto, el programa diseñado actúa diferente al anterior. Para unir estos datos, solo es necesario basarse en el nombre del pozo.

Se debió garantizar en la base de datos, que las bombas tengan el formato y nombre correcto, ya que, es necesario agregar las especificaciones técnicas de las bombas que dependen de los archivos creados en la etapa 1 (archivos derivados de los catálogos por marca). Para agregar estos datos, se realizó un sub-programa que permite al sistema buscar el nombre y marca de la bomba, extrayendo los datos de interés y colocándolos en la base de datos principal.

El diseño y creación de una herramienta de programación es complejo, por lo tanto, se debe tomar un solo pozo tipo, que permita mejorar y optimizar todos los procesos y programas realizados hasta lograr el resultado esperado para ser extrapolado a los otros pozos. En esta investigación se consideró como pozo tipo el pozo VD20, ya que, cumplía con la cantidad de datos mínimos requeridos, es uno de los pozos con mayores cambios de bomba y adicionalmente no posee ningún servicio de estimulación.

La base de datos principal se analizó de forma gráfica para cada pozo. Para esto, fue necesario desarrollar herramientas y funciones básicas de programación en Python, que permitieran filtrar los datos de forma rápida, por arena o por pozo, y generar una gráfica interactiva de dos dimensiones con doble eje Y. Este tipo de gráficas permitió comparar cada uno de los parámetros del historial de producción y cuál era su comportamiento con respecto a la tasa de petróleo.

En la figura 38, se observa un ejemplo de la gráfica interactiva programada. En el eje x, se muestra la línea de tiempo, en el eje Y principal (que se encuentra del lado izquierdo), se muestra la escala de la producción de petróleo y en el eje Y secundario (que se encuentra del lado derecho), se muestra la escala del parámetro a comparar. Adicionalmente, en la parte superior de la gráfica, se muestran en color naranja los cambios de bomba realizados a lo largo de la historia de producción, con la finalidad de verificar como afectan éstos en la tasa de petróleo. Si el pozo ha tenido algún procedimiento de estimulación deberá aparecer de igual forma como se muestran los cambios de bomba, pero, en otro color.

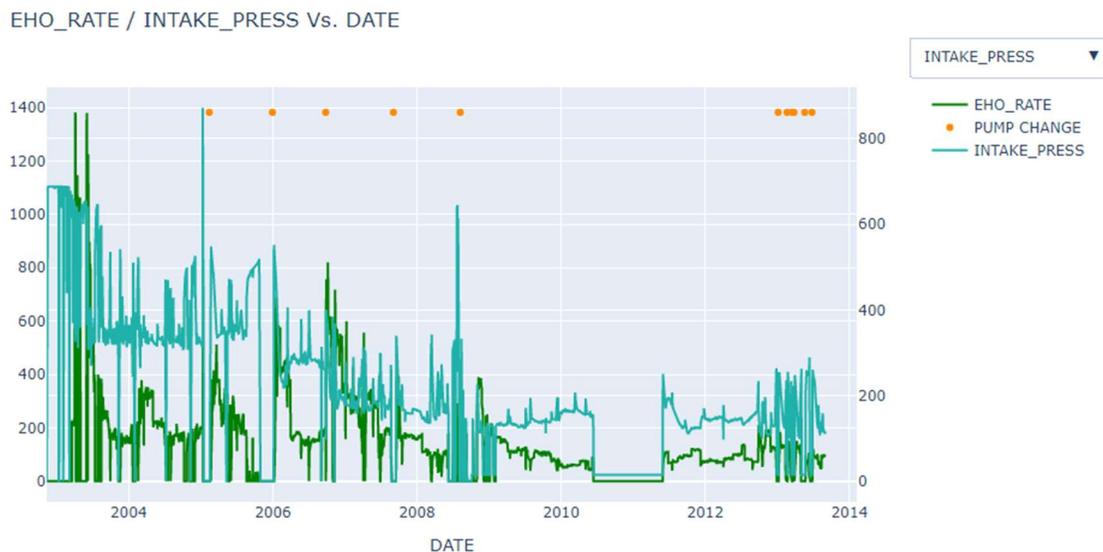


Figura 38: Gráfica interactiva para análisis de datos por parámetro en el pozo VD20

Este tipo de gráficos programados no solo permitió graficar y estudiar un solo parámetro versus la tasa de petróleo, sino que también permitió graficar varios parámetros y estudiar periodos de tiempos definidos a través de las herramientas de acercamiento. En la figura 39, se observa un claro ejemplo donde se comparan 4 parámetros como lo son las RPM, presión de entrada, caudal de agua y de petróleo del mismo pozo antes mostrado. Como se puede observar se estudia un periodo de tiempo definido que abarca desde mayo del 2006 a Julio 2007, y se muestra cómo se comparan los valores de cada uno de los parámetros de acuerdo con las etiquetas mostradas y su comportamiento con respecto a un día específico el cual está señalado en la parte inferior en un recuadro de color gris. Por otro lado, en la parte superior, se muestra la descripción del servicio más cercano a la fecha de estudio.



Figura 39: Gráfica interactiva para comparar y estudiar el comportamiento de más de un parámetro en un periodo de tiempo determinado del pozo VD20.

Este tipo de gráficas permitió verificar la calidad de las bases de datos. Las metodologías de aprendizaje automatizado se pueden aplicar a cualquier conjunto de datos, sin embargo, en las bases de datos se debe garantizar que no existan datos vacíos.

En las muestras analizadas se observó que existían datos erróneos como: datos de presión de fondo o temperatura por debajo de la atmosférica o presiones y temperaturas negativas, datos nulos o faltantes, eficiencias de la bomba mayores a uno, entre otros. Sin embargo, con la finalidad de determinar si estos datos tienen gran influencia en los resultados, se procedió a generar dos bases de datos, la primera es la base de datos original, sin ningún tipo de modificación, mientras que a la segunda base de datos se le aplicó caracterización estadística para eliminar los errores y la información faltante, este procedimiento se explica en la siguiente sección.

Por otra parte, se realizó un análisis de datos de las bombas en el campo, con la finalidad de estudiar su eficiencia tomando en cuenta el modelo y marca de la misma. Para este fin, fue necesario calcular la cantidad de días activos de la bomba. Este parámetro se calculó basado en el siguiente procedimiento: el primer paso consistió en generar una columna ("*Active_pump_days*") que determine a diario si la bomba esta activa, para esto se programó una función que toma en cuenta la tasa de petróleo y las RPM de la bomba, si ambos parámetros son mayores a cero entonces en la columna se registra como uno "1", el cual significa que la bomba esta activa, en cambio si no cumple la condición antes mencionada, entonces el programa coloca un cero "0", justificando que la bomba estaba inactiva. El segundo paso, fue realizar un ciclo, el cual recorra todos los días de producción y con respecto a la columna antes calculada, realice una sumatoria acumulativa de los días activos de la bomba. Este contador se reinicia una vez se dé un cambio de bomba en el pozo. En la figura 40, se observa el comportamiento de las bombas en el pozo tipo, donde cada periodo de curva en ascenso hasta el pico representa el tiempo activo de la bomba, y en la parte superior en color naranja se reflejan los cambios de bomba realizados.

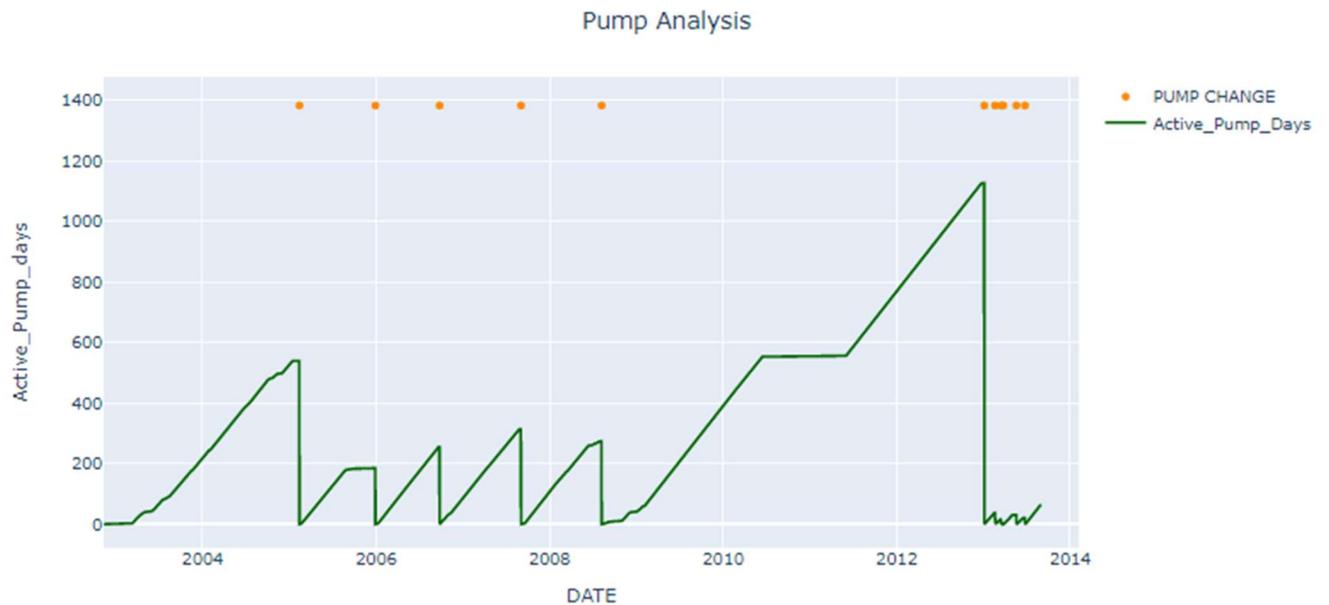


Figura 40: Análisis del comportamiento de las bombas presentes en el pozo VD20

Todo este análisis previo es necesario para proceder a la etapa 3, la cual comprende el análisis y preparación de los datos para posteriormente llevarlos a la etapa de modelado.

3.2.3.- Análisis y preparación de los datos

De acuerdo con el análisis realizado en la etapa anterior, se seleccionaron dos sets de datos. El primero corresponde a los datos originales, mientras el segundo corresponde a los datos base para el procedimiento de caracterización estadística, para la cual se diseñó un procedimiento el cual se explica a continuación:

3.2.3.1.- Análisis, identificación y visualización gráfica de datos perdidos

- Identificar el formato de los datos perdidos

El primer paso para la caracterización estadística fue identificar cuál es el formato de los datos perdidos. Existen dos tipos de datos perdidos, el primer tipo corresponde a los datos nulos, aquellos que no tienen ningún tipo de información y se caracterizan por encontrarse en celdas vacías o se representan con las siguientes nomenclaturas, N/A, o NaN. El segundo tipo de datos corresponde a datos erróneos, aquellos datos fuera de rango, negativos, o registrados con cero en parámetros o condiciones que no es posible obtener este valor. En la figura 41, se observa un esquema de los tipos de datos perdidos que se pueden encontrar en las bases de datos de estudio, sin embargo, pueden ser diferentes de acuerdo con los formatos en los que se trabaje, este paso es

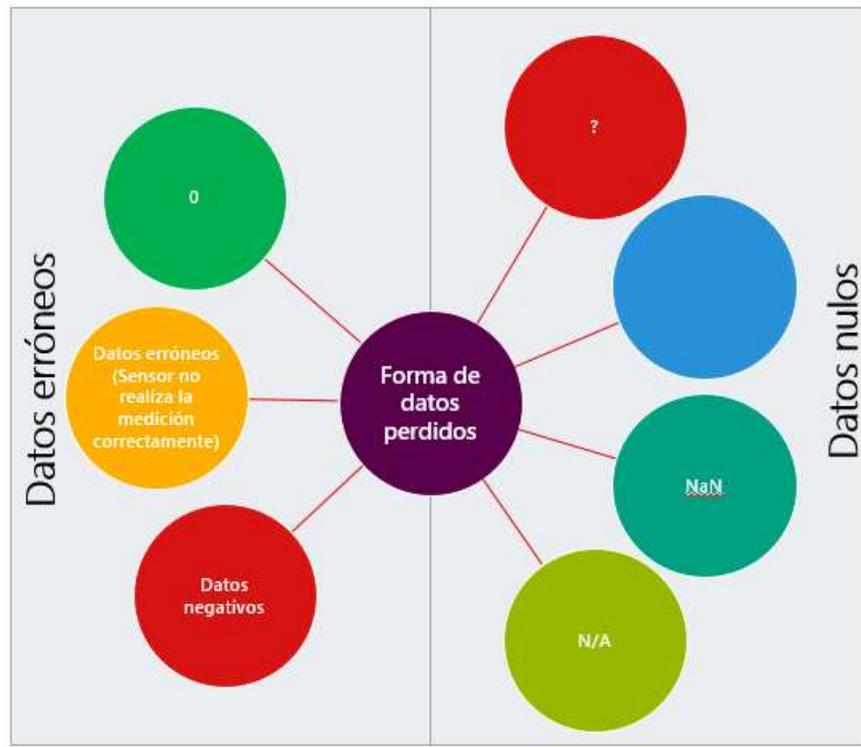


Figura 41: Esquema de datos perdidos

único en el procedimiento y todos los datos y nomenclaturas aquí registrados serán referencia para el análisis de todos los pozos de la muestra.

- Contabilizar datos nulos en el historial de producción

Una vez que se conoce el formato en el que se presentan los datos nulos, se procede a contabilizar la cantidad de datos nulos por columna en el historial de producción de cada uno de los pozos en estudio. Para lograr este procedimiento de forma rápida se programó una función algorítmica base, la cual, contabiliza la cantidad y el porcentaje de datos nulos con respecto al total de datos registrados del pozo. La figura 42, representa un ejemplo de la tabla de resultados del algoritmo programado en el pozo tipo. En ésta, se puede observar que los parámetros resaltados en amarillo presentan datos nulos, sin embargo, el porcentaje de los mismos no es alto, por el contrario, aquellos parámetros resaltados en color rojo presentan un porcentaje muy alto de datos nulos, por lo tanto, es necesario eliminar esta columna para el entrenamiento, ya que los datos no tienen relevancia y no podrán ser completados en el estudio.

	Cantidad de Datos Nulos	Porcentaje de Datos nulos %		Cantidad de Datos Nulos	Porcentaje de Datos nulos %
WELLNAME	0	0.000000	PI	14	0.354790
DATE	0	0.000000	PUMP_EFFICIENC	0	0.000000
DURATION	0	0.000000	EHO_RATE	0	0.000000
WHP	2	0.050684	CASING_PRESS	3458	87.633046
WHT	2	0.050684	CLUSTER	0	0.000000
RPM	0	0.000000	TRUNKLINE	0	0.000000
TORQUE	1	0.025342	RESERVOIR	0	0.000000
INTAKE_PRESS	0	0.000000	ASSET	0	0.000000
DISCHARGE_PRES	2	0.050684	YEAR	0	0.000000
DISCHARGE_TEMP	1	0.025342	MONTH	0	0.000000
DILUENT_VOL	0	0.000000	GOR	0	0.000000
GAS_RATE	0	0.000000	WOR	0	0.000000
WATER_RATE	0	0.000000			

Figura 42: Ejemplo de resultados al contabilizar los datos nulos en el pozo VD20

- Identificar las fechas donde se encuentran los datos nulos

Basados en el paso anterior se procedió a programar un algoritmo que permitió identificar y filtrar de forma rápida las fechas exactas en las que están presentes estos datos nulos en cada uno de los pozos con la finalidad de poder aplicar la caracterización estadística con detalle. En la figura 43, se observa un ejemplo claro de un pozo con datos nulos en el historial de producción, y como se pueden identificar. Es importante tomar en cuenta que la existencia de los datos no garantiza que sean correctos, es por eso, que se deben estudiar los datos erróneos.

WELLNAME	DATE	DURATION	WHP	WHT	RPM	TORQUE	INTAKE_PRESS	DISCHARGE_PRES	DISCHARGE_TEMP	DILUENT_VOL	
0	VD20	2003-02-04	0.000	NaN	NaN	0.000	NaN	0.000000	NaN	NaN	0.000000
1	VD20	2003-09-05	24.000	NaN	NaN	89.049	885.687837	332.373169	1005.596982	123.706667	180.489231
2	VD20	2006-04-12	23.617	213.242792	109.535	240.987	533.815003	295.069344	NaN	138.439400	0.000000

PI NULL DATA(14)	WHP (2) WHT (2) NULL DATA	DISCHARGE_PRES (2) NULL DATA	TORQUE (1) DISCHARGE_TEMP (1) NULL DATA
Período	2003-02-04	2003-02-04	2003-02-04
2009-02-13 / 2009-02-26	2003-09-05	2006-04-12	2003-02-04

Figura 43: Ejemplo de datos nulos presentes en el historial de producción del pozo VD20

- Estudio de los rangos de datos de cada uno de los atributos

Como se explica en el paso anterior, la existencia de los datos no garantiza que éstos sean correctos o tengan sentido físico, por lo tanto, se debe ser cuidadoso con las bases de datos antes de proceder a aplicar algoritmos de ML. El estudio de los rangos está asociado a conocimiento técnico y conocer con detalle el proceso de adquisición de datos en el campo, con la finalidad de definir los rangos aceptables y con sentido físico. Adicionalmente se requiere conocimiento estadístico como media (mean), desviación estándar (std), valor mínimo (min), valor máximo (max) y percentiles (25%, 50%, 75%, según sea el caso), para identificar la calidad de los datos. Existen

parámetros como las presiones y temperaturas medidas en el fondo, las cuales no tienen un máximo definido, ya que dependen de la presión y temperatura del yacimiento, al igual que las presiones y temperaturas en superficie, las cuales deberían ser iguales a la atmosférica siempre y cuando no estén bajo una perturbación. Por esta razón, es importante analizar con detalle el periodo inicial de producción. Para esto, se programó una función algorítmica base la cual identifica el periodo inicial de cada pozo y grafica cada uno de los atributos con respecto a la tasa de producción de petróleo, con la finalidad de garantizar que no exista producción en este periodo. Este algoritmo se basa en identificar la fecha de inicio de producción del pozo y contabilizar la cantidad de días con producción cero. Una vez la función selecciona el periodo inicial, se estudian los valores de los parámetros de interés antes mencionados y con estos datos se puede identificar los valores máximos y mínimos referenciales para cada atributo. En la figura 44, se observa un ejemplo del periodo inicial del pozo tipo para analizar el parámetro INTAKE_PRESS. En el lado izquierdo se observa el gráfico antes mencionado y en el lado derecho se observan todos los cálculos estadísticos donde se obtendrán los valores necesarios para definir los rangos.

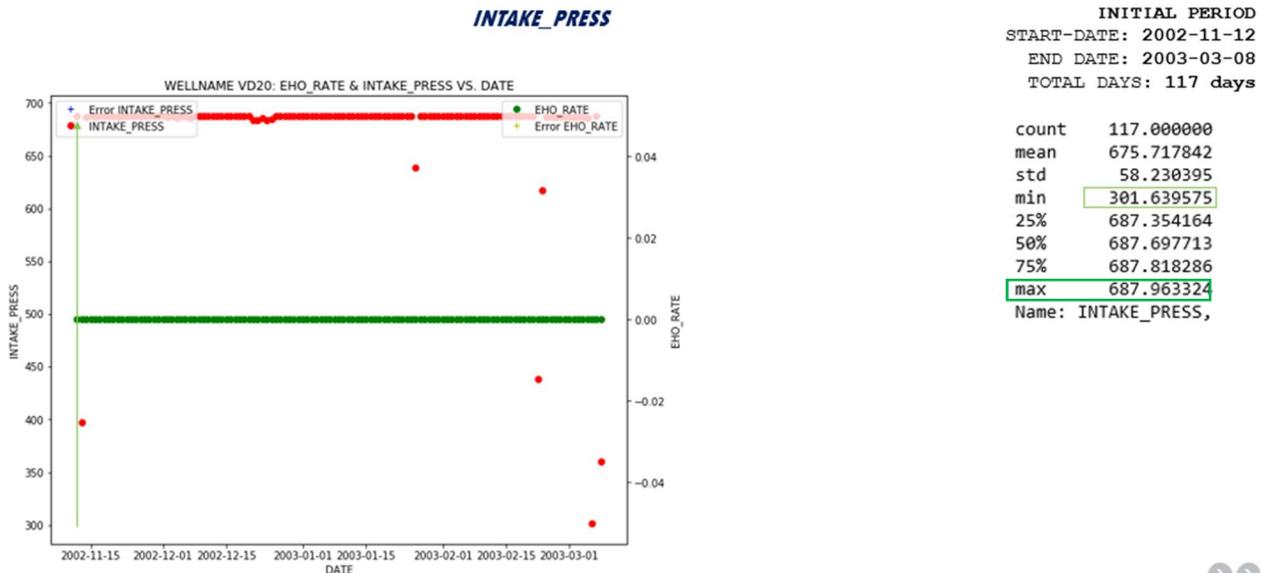


Figura 44: Estudio del periodo inicial del pozo VD20

- Identificar y graficar todas las anomalías y servicios de interés en el historial de producción

Una vez se determinaron los rangos aceptables para cada uno de los parámetros se procedió a identificar las fechas de los datos fuera de rango en el historial de producción y crear una tabla completa con todas las anomalías incluyendo los datos nulos ya identificados, al igual que los servicios de interés como, cambios de bomba o estimulaciones. Para este fin se programó una nueva función algorítmica base que muestre un cuadro estadístico de todo el historial de producción del pozo y permita graficar las anomalías (datos erróneos y nulos) con un color particular, al igual que los servicios de interés. En la figura 45, se observa un ejemplo gráfico del pozo tipo sin servicios de estimulación, en donde se aprecian los datos erróneos con puntos dispersos en color vinotinto y en color negro en la parte superior los datos nulos. Adicionalmente se observan los cambios de bomba para un mejor análisis. Terminado el proceso de identificación, se inició el proceso de caracterización estadística.

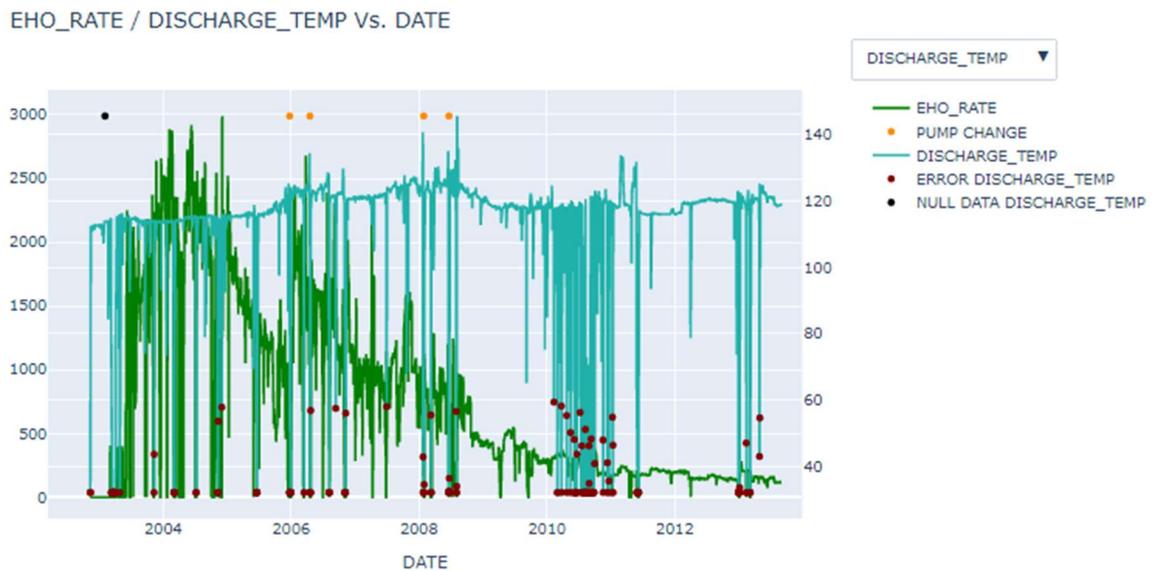


Figura 45: Ejemplo de estudio de anomalías en el historial de producción en el pozo VD20

3.2.3.2.- Metodologías para la caracterización estadística

La caracterización estadística es un procedimiento que forma parte del *Data Mining*, el cual consistió en rellenar los datos de acuerdo a valores estadísticos, con la finalidad de obtener datos con sentido físico y sin vacíos de información. En este proceso es necesario crear distintos algoritmos básicos que permitan modificar los datos, basados en el comportamiento de cada uno de los parámetros, para lograr una buena caracterización estadística. Es importante tomar en cuenta que estos arreglos generan una aproximación y que no son los valores reales, por lo tanto, generará un porcentaje de error en los resultados. Para esta metodología se programaron diez funciones diferentes, valor mínimo (*minimum value*), valor máximo (*maximum value*), valor medio (*middle value*), valor promedio (*mean value*), corrección lineal basado en dos puntos (*linear correction*), regresión lineal (*linear regression*), corrección de la presión de entrada (*correct intake press*), restauración de presión (*build up*), valores constantes (*constant values*) y reemplazar valores específicos (*replace value*) las cuales se explican con detalle a continuación:

- Valor mínimo (*minimum value*): para aplicar este método es necesario definir dos periodos, un periodo de entrenamiento y un periodo de aplicación. El periodo de entrenamiento es definido por el usuario considerando el atributo a modificar y otros atributos de referencia que sean similares en ambos periodos. Este algoritmo toma el valor mínimo del periodo de entrenamiento y sustituye todos los valores del mismo atributo en el periodo de aplicación.
- Valor máximo (*maximum value*): es similar al de valor mínimo, a diferencia de este, el algoritmo extrae el valor máximo del periodo de entrenamiento y modifica todos los datos del periodo de aplicación con este dato.
- Valor medio (*middle value*): aplicable a valores nulos o fuera de los rangos especificados. Está diseñado para modificar un solo dato en una fecha definida por el usuario. Ésta función toma como referencia el dato del atributo el día

anterior y el valor posterior a la fecha definida, con ambos datos calcula un promedio y el resultado se utiliza para rellenar el dato faltante. En la figura 46, se observa la aplicación de este método en el parámetro de presión de entrada de la bomba (*Intake_Press*) del pozo tipo, en la figura 46a, se observa una caída drástica en el valor de la presión de entrada la cual se resalta en el recuadro de color rojo, mientras en la figura 46b, se observa el parámetro ya caracterizado en este punto, el cual se resalta en el recuadro de color verde.

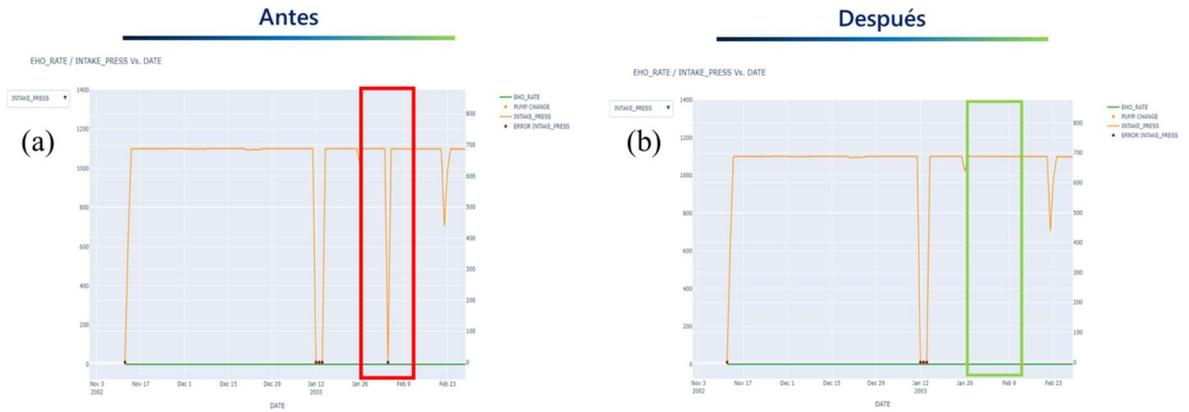


Figura 46: Ejemplo de caracterización estadística por valor medio en el pozo VD20
 (a) Parámetro sin caracterizar (b) Parámetro caracterizado

- Valor promedio (*mean value*): requiere dos periodos, uno de entrenamiento y otro de aplicación. Una vez que el usuario define el periodo de entrenamiento, el algoritmo calcula el promedio de los datos seleccionados y sustituye todos los datos del atributo de estudio en el periodo de aplicación por el valor promedio obtenido. En la figura 47, se observa un ejemplo de este método aplicado en el pozo tipo para el parámetro de temperatura de cabezal. En la figura 47a, se observa el periodo de entrenamiento, donde las condiciones de producción de petróleo son muy similares al de periodo de aplicación, por lo tanto, el algoritmo calcula el promedio de temperatura en el periodo de entrenamiento y lo sustituye en el periodo de aplicación. En la figura 47b, se observa el periodo de aplicación sin realizar el cambio, la cual se representa con un cuadro de color rojo, mientras en la figura 47c, se observa el parámetro ya

modificado en el periodo seleccionado, el cual se representa con el cuadro de color verde.

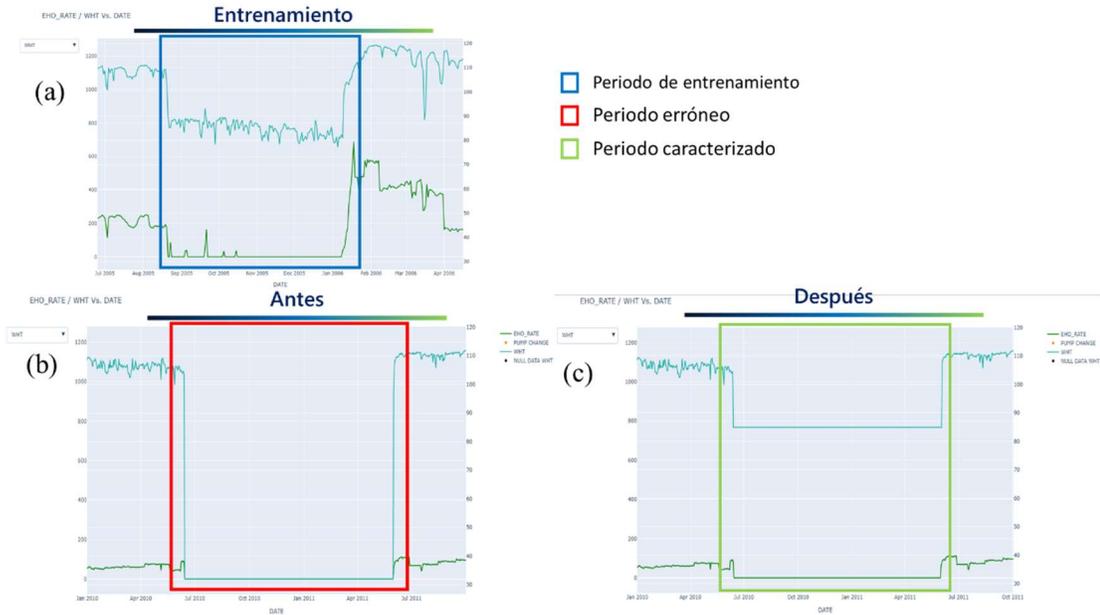


Figura 47: Ejemplo de caracterización estadística por valor promedio en el pozo VD20 (a) Parámetro sin caracterizar (b) Parámetro caracterizado

- Corrección lineal basado en dos puntos (*linear correction*): para este método se selecciona el periodo con datos erróneos incluyendo un punto de inicio y un punto final, los cuales deben tener datos correctos del atributo a completar. Con estos dos datos el algoritmo permite determinar la ecuación de la recta entre dos puntos y con esta completa los datos faltantes en el periodo seleccionado.
- Regresión lineal (*linear regression*): se basa en una metodología de *machine learning*, donde el algoritmo usa los datos de dos atributos seleccionados por el usuario y encuentra la ecuación lineal que más se ajusta a ambos parámetros. Para aplicar este método es necesario seleccionar dos periodos, uno de entrenamiento y otro de aplicación, el usuario debe verificar que los dos parámetros tienen relación lineal en el periodo de entrenamiento seleccionado, una vez que el algoritmo determina la ecuación de la recta característica en dicho periodo, aplica esta misma ecuación y calcula los datos faltantes o

erróneos en el periodo de aplicación. Para este método existe un caso particular el cual está asociado a la presión de entrada de la bomba (*Intake_press*) respecto a la producción de petróleo, ya que, en la mayoría de los casos estos estaban siempre inversamente relacionados, por esta razón, se genera un método llamado corrección de la presión de entrada (*correct intake press*). En la figura 48a, se observa el periodo de entrenamiento, donde la medición de la presión de entrada mantiene el comportamiento respecto a la tasa de petróleo, e incluso el parámetro de producción de petróleo es similar en ambos periodos, tanto el de entrenamiento como el de aplicación. En la figura 48b, se observa el error en la medición de la presión de entrada en el recuadro de color rojo, mientras en la figura 48c, se observa la presión ya modificada por el método de regresión lineal, la cual se representa en el recuadro de color verde.

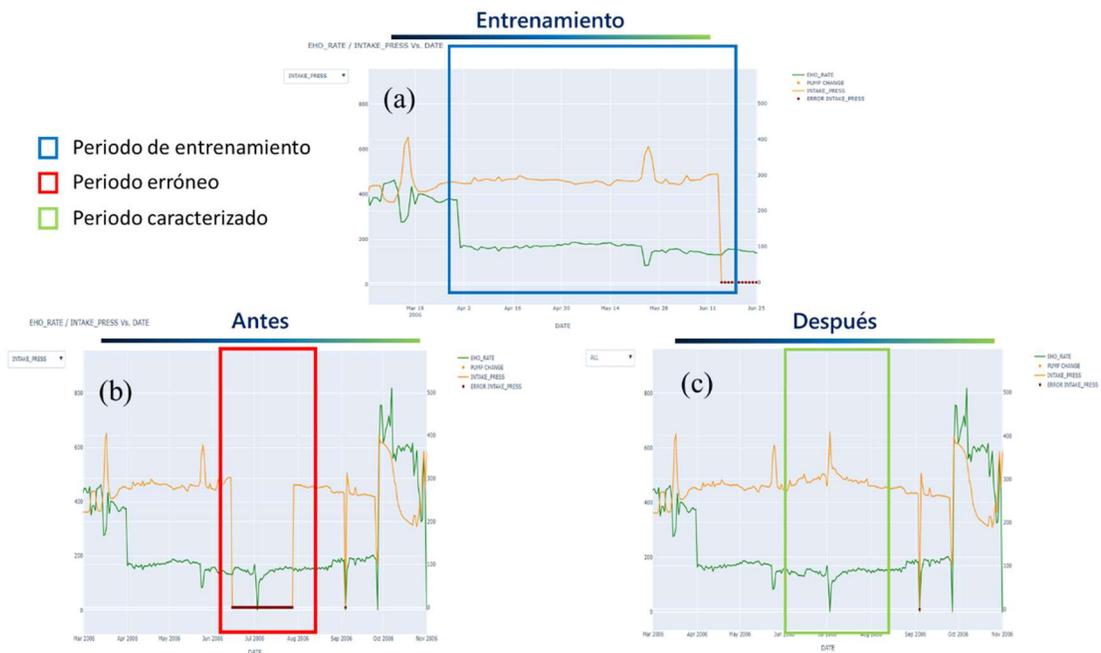


Figura 48: Ejemplo de caracterización estadística por regresión lineal en el pozo VD20 (a) Parámetro de presión de entrada sin caracterizar (b) Parámetro de presión caracterizado

- Restauración de presión (*build up*): diseñado sólo para la modificación de presiones a nivel de subsuelo, se basa en otra metodología de ajuste de *machine*

learning en la cual el usuario define una ecuación característica, en este caso la ecuación suministrada al sistema es la ecuación de las pruebas de restauración de presión, con la finalidad de que el algoritmo ajuste los datos de forma logarítmica, halle los parámetros desconocidos de ajuste y su comportamiento sea similar a este tipo de curvas. En este método el usuario debe seleccionar dos periodos, el periodo de entrenamiento y el periodo de aplicación. El periodo de entrenamiento debe cumplir con dos condiciones, la primera condición es que la tasa de producción en el periodo seleccionado debe ser igual a cero, por otro lado, el comportamiento de la curva debe cumplir con la tendencia logarítmica, mientras el periodo de aplicación debe cumplir con la primera condición de tasa igual a cero e incluir dos datos importantes: en la fecha de inicio debe tener la última presión real medida durante el periodo de producción anterior y en la fecha final debe tener la presión promedio del pozo en estudio. Es importante tomar en cuenta que esta función está programada para ajustarse a las distintas características que presentan los pozos. En la figura 49a, se observa el periodo de entrenamiento del pozo, donde existe producción cero y se ve de forma clara la tendencia de restauración de presión, mientras en la figura 49b se observa el periodo de aplicación donde la presión cae drásticamente a cero al caer la producción, por lo tanto, al aplicar la metodología se aprecia como la presión sigue el patrón del periodo de entrenamiento como se observa en la figura 49c.

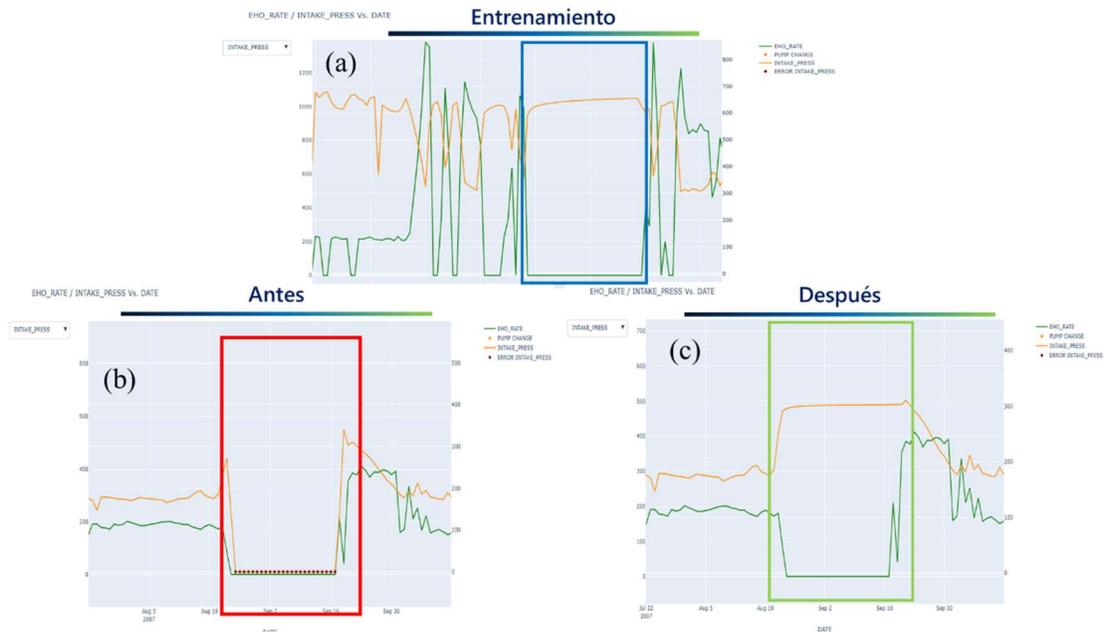


Figura 49: Ejemplo de caracterización estadística por el método de restauración de presión en el pozo VD20 (a) Parámetro de presión de entrada sin caracterizar (b) Parámetro de presión caracterizado

- Valores constantes (*constant values*): se debe definir un periodo de aplicación y un valor específico, en el cual los datos en el periodo seleccionado se modifican y toman como dato el valor específico.
- Reemplazar valores específicos (*replace value*): último método programado, en el cual se define un periodo y dos valores específicos, el primer valor corresponde al valor a modificar dentro del periodo seleccionado, mientras el segundo valor corresponde al dato que va a modificar al primero dentro del periodo definido.

Todos estos métodos fueron necesarios para lograr la caracterización estadística y una buena calidad de los datos en el segundo set de datos.

Al finalizar el procedimiento de caracterización estadística, se aplicó la metodología para la transformación de datos a *dummies* en ambos sets de datos (originales y caracterizados).

3.2.3.3.- Transformación de datos de texto a *Dummy data*

Este procedimiento consistió en transformar los datos de tipo texto en variables numéricas o de código binario. En el caso de este TEG, es necesario entender cuál es el comportamiento de los modelos de bombas en los pozos, al igual que la respuesta que tiene un proceso de estimulación del yacimiento en la producción de un pozo, por lo cual, es necesario realizar dos procedimientos: el primero que permita transformar los datos de los modelos de las bomba en datos numéricos, y el segundo que transforme los trabajos de estimulación en una variable numérica que pueda ser entendida por el algoritmo.

- *Dummy data* para los modelos de BCP: con la finalidad de transformar los datos de los modelos de bombas en datos numéricos, este algoritmo permite crear columnas por cada uno de los modelos presentes en la columna ‘*Pump_model*’ eliminando la misma de la base datos. La cantidad de columnas resultantes varia, ya que se crean varios parámetros adicionales dependiendo de la cantidad de modelos de bombas que se usan a lo largo de la vida productiva del pozo. Una vez que se crean estas columnas, el algoritmo trabaja como una función de activación que recorre día a día la base de datos y de acuerdo con la bomba que esté activa, este coloca un “1” en la columna de esa bomba, mientras las otras columnas de los otros modelos se declaran inactivas con un “0”. Este tipo de funciones facilita el estudio de estos parámetros de forma matemática y estadística.
- *Dummy data* para trabajos de estimulación: las estimulaciones son realizadas en algunos pozos, por lo tanto, el algoritmo permite crear una columna adicional llamada “*Stimulation_effect*” la cual actúa muy similar al *Dummy data* de las

BCP, sin embargo, solo se crea esta única columna. Inicialmente esta columna se caracteriza por ser una columna de ceros “0”, la cual indica que el efecto de estimulación se encuentra inactivo. Sin embargo, el algoritmo tiene la capacidad de identificar cuando el pozo ha sido expuesto a un trabajo de estimulación, por lo cual, a partir del día del trabajo esta columna se declara activa, colocando un “1” en la misma, por los siguientes 90 días. Este tiempo es considerado un tiempo promedio en el cual tiene efecto la estimulación en el campo. Este tipo de variables permite definir si el trabajo de estimulación realmente tiene un efecto positivo sobre la producción del pozo.

La transformación de datos, junto a la caracterización estadística son fundamentales para el proceso de predicción, ya que garantizan el cumplimiento de los objetivos y una mejora en la calidad del dato.

3.2.3.4.- Creación de *datasets* finales para el modelado

Este paso es el último en el análisis y preparación de los datos, el cual busca generar los sets de datos finales para ingresar a la fase de modelado. Con la finalidad de estudiar el tiempo de cómputo y determinar el set de datos óptimo en la fase de modelado se generaron tres tipos de *datasets* principales para este TEG: el primer tipo de *dataset* corresponde a los datos reales, el segundo tipo corresponde a los datos normalizados y por último el tercer tipo de *dataset* que corresponde a los datos estandarizados.

- Set de datos reales

Este conjunto de datos es único para cada pozo y trabaja con los datos obtenidos de la medición en campo. Sin embargo, con la finalidad de estudiar la importancia de la calidad del dato se generaron dos conjuntos de datos reales, el primer conjunto

corresponde a los datos crudos (*raw data*) que consiste en los datos cuyos valores no han sido alterados de ninguna manera (sin procesar con la caracterización estadística), con la excepción de aquellos datos que estuviesen vacíos, que fueron eliminados de la muestra para evitar errores durante el entrenamiento, mientras el segundo conjunto corresponde a los datos reales caracterizados previamente.

- Set de datos normalizados

Este conjunto de datos es un derivado de los datos reales a los cuales se les aplicó un proceso de normalización, el cual permitió escalar cada uno de los parámetros en estudio con un valor de distribución entre 0 y 1, el cual se calcula con la ecuación 9.

Este procedimiento es aplicado a ambos *datasets*, tanto al set de datos sin procesar como al set de datos caracterizados, dando como resultado dos nuevos *datasets* con valores que van entre cero y uno en todas sus variables numéricas.

- Set de datos estandarizados

Este conjunto de datos también es un derivado de los dos dataset reales, a los cuales se les aplicó la metodología de estandarización, los *datasets* son reescalados para asegurar que la media y la desviación estándar sean 0 y 1, respectivamente. Este procedimiento es útil para los algoritmos de optimización, como el descenso de gradiente que se utiliza en los algoritmos de aprendizaje automático, que ponderan las entradas (por ejemplo, las redes neuronales y de regresión). Este procedimiento permitió obtener como resultado dos nuevos *datasets*, datos estandarizados sin caracterizar y datos estandarizados y caracterizados.

Una vez obtenidos los tres tipos de *datasets* para ambos conjuntos de datos (reales sin caracterizar y caracterizados), fue necesario programar una función que permitiera separar cada *dataset* en dos, un set de datos de entrenamiento (*training data*) y un set

de datos de prueba (*test data*). Para separarlos hay que tener en cuenta varios factores importantes como: el porcentaje de datos que se van a usar para cada caso y si los datos son dependientes del tiempo.

Usualmente en los métodos de aprendizaje supervisado de *machine learning*, el 70% de los datos son usados para el entrenamiento, mientras que el otro 30% corresponden a los datos de prueba, por lo tanto, se mantuvo el porcentaje para este TEG. Sin embargo, en la mayoría de los casos, ambos sets de datos se toman al azar, y el programa va variando los datos en cada iteración. Éste, realiza un gran número de iteraciones con la finalidad de encontrar el mejor ajuste, por lo tanto, a medida que aumenta el número de iteraciones aumenta el número de datos analizados. En este caso, como las variables dependen del tiempo, fue necesario tomar los datos de manera cronológica, con la finalidad de estudiar el comportamiento de las bombas a lo largo de todo el desarrollo del pozo, y adicionalmente estudiar las consecuencias del desgaste mecánico en cada bomba y la declinación del yacimiento, por esta razón, no existe variación en los datos de entrada y solo se realiza una única iteración. Para este procedimiento se realizó un algoritmo el cual genera diferentes sets de datos por cada una de las bombas presentes en el pozo, tomando en cuenta la fecha inicial y la fecha final de su actividad. Cada uno de estos sets de datos fue separado a su vez en dos nuevos *datasets*, un *dataset* de entrenamiento con un 70% de los datos desde su fecha de inicio y consecutivos, y otro *dataset* de prueba que tomó el 30% restante. En la figura 50, se observa el historial de producción de petróleo de un pozo en color verde, mientras en líneas rojas se representan los cambios de las bombas a lo largo de su producción, en la parte superior de cada periodo entre los cambios de bomba está representado en color naranja el 70% de los datos correspondientes al entrenamiento, y en color azul los datos correspondientes a los datos de prueba. Una vez se realizó este procedimiento para cada bomba, se concatenan todos los *datasets* de entrenamiento en uno solo y los *datasets* de prueba en otro. Éste procedimiento es realizado para los seis *datasets* principales de cada pozo (reales corregidos, reales sin caracterizar y sus derivados, normalizados y estandarizados).

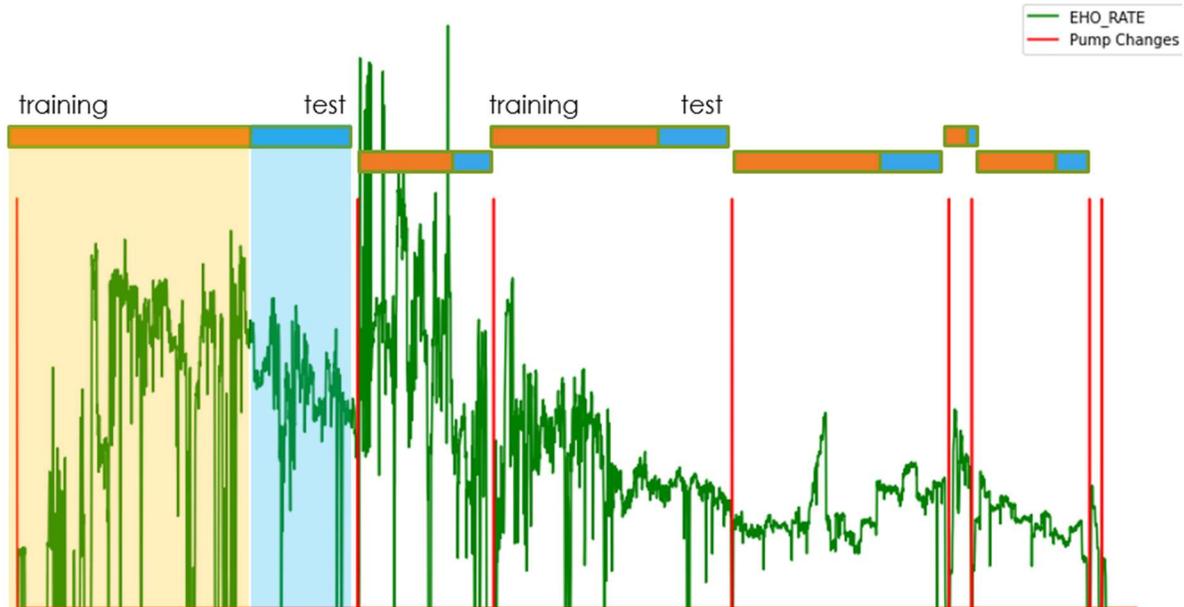


Figura 50: Ejemplo de separación de datasets de entrenamiento y prueba

- Al finalizar el análisis y preparación de los datos se procedió a seleccionar las técnicas y metodologías de *Data Science* a desarrollar en la etapa de modelado. Esta etapa busca describir la tendencia a lo largo del tiempo de los parámetros de un grupo de pozos en estudio. De acuerdo a lo explicado en el capítulo II, los métodos de predicción a utilizar en este TEG se basan en dos disciplinas, *machine learning*, en la cual se desarrollaron 4 metodologías: dos metodologías para la predicción de producción (regresión lineal multivariable y la regresión polinómica) y dos metodologías para el análisis de bombas (árbol de decisiones y KNN), mientras que para la disciplina de *deep learning*, el método de predicción de la producción se basa en redes neuronales.

Es importante tomar en cuenta que antes de la fase de modelado, todos los datos deben pasar por la etapa de preprocesamiento, con la finalidad de optimizar y facilitar los cálculos en los algoritmos de modelado, los cuales se explican en el siguiente apartado.

3.2.4.- Modelado y evaluación

El modelado de esta tesis se divide en dos partes, un primer modelado y evaluación de la producción y un segundo modelado aplicado al análisis de bombas.

3.2.4.1.- Modelado de la producción

Para la aplicación de todos los métodos de predicción expuestos en este TEG fué necesario desarrollar un algoritmo que permitiera identificar las variables numéricas y de tipo *Dummy* de interés para el estudio en la base de datos principal, y que adicionalmente sean aquellas variables que pudiesen influir en la producción de petróleo que es el dato que se busca predecir. En la tabla 7, se explica cada uno de los parámetros en la base de datos y el formato en el que están presentes, mientras, en la tercera columna llamada uso, se especifica si el dato será utilizado para el aprendizaje del algoritmo. Por último, se especifica el tipo de parámetro, el cual define si el parámetro será utilizado como parámetro de entrada o de salida, esto puede variar de acuerdo a cada metodología.

Tabla 7: Tabla de atributos finales para el modelado (Caso base)

Dato	Tipo	Uso	Tipo de Parametro
<i>WELLNAME</i>	Texto	NO (Clasificación)	
<i>DATE</i>	Fecha	NO	
<i>DURATION</i>	Numérico	SI	Entrada
<i>WHP</i>	Numérico	SI	Entrada
<i>WHT</i>	Numérico	SI	Entrada
<i>RPM</i>	Numérico	SI	Entrada
<i>TORQUE</i>	Numérico	SI	Entrada
<i>INTAKE_PRESS</i>	Numérico	SI	Entrada
<i>DISCHARGE_TEMP</i>	Numérico	SI	Entrada
<i>DILUENT_VOL</i>	Numérico	SI	Entrada
<i>GAS_RATE</i>	Numérico	SI	Entrada
<i>WATER_RATE</i>	Numérico	SI	Entrada
<i>PI</i>	Numérico	SI	Entrada
<i>PUMP_EFFICIENC</i>	Numérico	SI	Entrada
<i>ALLOCPRUEBA_DATE</i>	Fecha	NO	
<i>STATUS</i>	Texto	NO	
<i>EHO_RATE</i>	Numérico	SI	Salida
<i>CLUSTER</i>	Texto	NO (Clasificación)	
<i>TRUNKLINE</i>	Texto	NO (Clasificación)	
<i>RESERVOIR</i>	Texto	NO (Clasificación)	
<i>ASSET</i>	Texto	NO (Clasificación)	
<i>T</i>	Numérico	SI	Entrada
<i>GOR</i>	Numérico	Implicito (Parametro calculado)	
<i>WOR</i>	Numérico	Implicito (Parametro calculado)	
<i>LIQUID_RATE</i>	Numérico	Implicito (Parametro calculado)	
<i>LEGAL_WELLNAME</i>	Texto	**	
<i>Event_Type</i>	Texto	Implicito	
<i>Event_Objective</i>	Texto	**	

Tabla 7: Tabla de atributos finales para el modelado (Caso base)(Continuación)

<i>Start_Date_Of_Event</i>	Fecha	NO	
<i>End_Date_Of_Event</i>	Fecha	NO	
<i>Pump_Model</i>	<i>Dummy</i>	SI	Entrada
<i>Pump_brand</i>	Texto	Implicito	
<i>Stimulation_effect</i>	<i>Dummy</i>	Si	
<i>TMD</i>	Numérico	*	
<i>Sensor_TMD</i>	Numérico	*	
<i>Sensor_TVD</i>	Numérico	*	
<i>Category</i>	Numérico	NO (Clasificación)	
<i>Category_name</i>	Texto	Implicito	
<p>* Descartados para la investigación por ser un valor fijo respecto al tiempo ** Descartados por no formar parte de los atributos de interés Implicitos: Valores calculados o clasificados en base a otros parámetros</p>			

Conocer la tabla de datos, el significado de cada parámetro y el lugar de su medición fue fundamental para el desarrollo de los algoritmos de aprendizaje automatizado.

- Modelo de regresión lineal multivariable

Esta metodología se explica con detalle en el capítulo II. Inicialmente se estudiaron los 13 parámetros de entrada de tipo numérico, sin embargo, los parámetros de entrada pueden variar de acuerdo con la variable *Dummy* (“*Pump_Model*”), la cual como se explicó antes aumentará de acuerdo a la cantidad de bombas existentes en el pozo a estudiar. Adicionalmente, para aplicar este método se requiere 1 parámetro de salida, que para esta investigación es la tasa de petróleo (“*EHO_RATE*”).

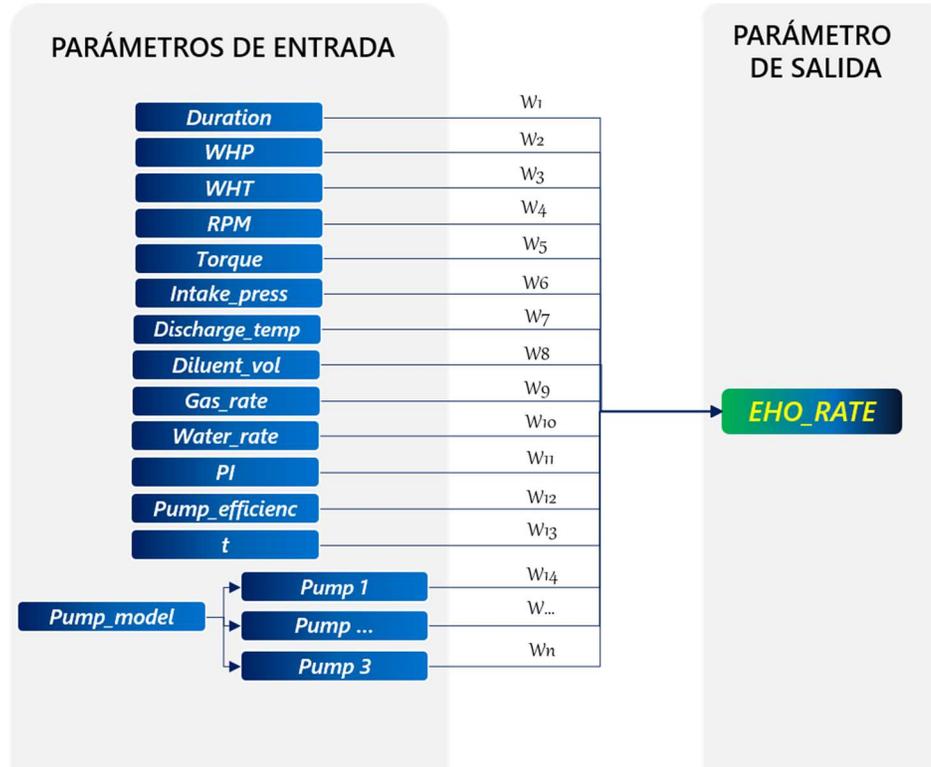
Una vez que los datos para el aprendizaje fueron definidos, fue necesario estudiar los 6 casos como se muestran en la tabla 8: el primer caso se basa en el entrenamiento con los datos reales sin caracterizar, el segundo caso con los datos reales caracterizados, el tercer caso corresponde a los datos reales sin caracterizar normalizados, el cuarto caso a los datos caracterizados y normalizados, el quinto caso

es con datos sin caracterizar pero estandarizados y por último el caso con datos caracterizados y estandarizados. En la mayoría de los estudios de *Data science* los datos se estudian tanto normalizados como estandarizados para disminuir el tiempo de cómputo.

Tabla 8: Casos según el set de datos a utilizar

Set de datos			
Tipos	Reale	Normalizado	Estandarizado
Originales	Caso 1	Caso 3	Caso 5
Caracterizados	Caso 2	Caso 4	Caso 6

El algoritmo está programado para trabajar inicialmente con los datos de entrenamiento de cada caso, con la finalidad de encontrar un patrón representativo y comparar los resultados obtenidos con los datos de prueba. Éste, estudia la relación lineal que tiene cada uno de los atributos de entrada con respecto al atributo de salida, calculando los pesos o coeficientes que multiplican a cada parámetro para que el resultado de la suma de todos los atributos multiplicados por su respectivo coeficiente se acerque lo más posible al parámetro de salida como se mostró en la figura 22 del capítulo II. En esta metodología los coeficientes presentan una correspondencia 1:1 respecto a los parámetros de entrada, lo que quiere decir que para cada parámetro de entrada le corresponde un único coeficiente. En la figura 51, se muestra un esquema con los parámetros de entrada iniciales del estudio para la regresión lineal multivariable y en la parte inferior la ecuación característica del modelo. Este esquema será la base de todas las pruebas que se realicen con la finalidad de encontrar el esquema óptimo.



$$EHO_RATE = (W1 * Duration) + (W2 * WHP) + (W3 * WHT) + (W4 * RPM) + (W5 * Torque) + (W6 * Intake_press) + (W7 * Discharge_press) + (W8 * Diluent_vol) + (W9 * Gas_rate) + (W10 * Water_rate) + (W11 * PI) + (W12 * Pump_efficienc) + (W13 * Casing_press) + (W14 * Pump1) + \dots + (Wn * Pumpn) + Constante$$

Figura 51: Esquema base del modelo de regresión lineal multivariable

El algoritmo desarrollado tiene la capacidad de determinar la ecuación característica. Con esta ecuación el algoritmo calculó la tasa de petróleo de la predicción para todos los datos, incluyendo set de entrenamiento y de prueba. Con los resultados obtenidos de la predicción, se calcularon una serie de errores y parámetros necesarios para estudiar la eficiencia del modelo, los cuales se muestran en la tabla 9:

Tabla 9: Parámetros calculados para estudiar eficiencia del modelo

Errores y Parametros
Tiempo (s)
Total de datos analizados
r ²
Error relativo promedio (%)
Error absoluto promedio (bpd)
Error promedio de la producción sobreestimada(bpd)
Error promedio de la producción subestimada(bpd)
Tasa total de petróleo (bpd)
Tasa de petróleo estimada (bpd)
Error entre la tasa total estimada y la tasa total real (%)

Tiempo (s): indica el tiempo en el que modelo logró encontrar la ecuación representativa del modelo y es calculado a través de una función interna dentro del mismo.

Total de datos analizados: indica la cantidad total de datos estudiados por el modelo en cada una de las iteraciones. En este caso por representar un patrón dependiente del tiempo, sólo se estudia con una sola iteración para el entrenamiento.

R cuadrado: indica el valor de ajuste entre las variables en estudio, entre más cercano 1 se encuentre, demuestra que existe mayor relación, y es calculado directamente por la función programada.

Error relativo promedio (%): indica la diferencia entre el dato real y el dato calculado por la predicción respecto al real y es expresado en porcentaje. Es muy común que al momento de calcularlo el resultado arroje algún error, ya que en el procedimiento se debe realizar la división entre los datos reales. Si en algún día del historial de producción éste tiene una producción de petróleo igual a cero, entonces el valor del error en ese punto es indeterminado o infinito. Con la finalidad de evitar este error, se ubicaron aquellas tasas donde la tasa de petróleo real o las RPM de la bomba eran igual a cero, y el programa inmediatamente asume que su predicción, será igual a cero, forzando así algunos resultados en la predicción, por lo tanto, estos datos se omitieron para encontrar el error relativo promedio final, el cual se calcula de la sumatoria diaria del error relativo dividido entre la cantidad de días en estudio.

Error absoluto promedio (bpd): este error determina la diferencia entre la cantidad de barriles reales y la predicción. En la figura 52, se muestra un ejemplo de la predicción generada por el algoritmo, en la cual la línea color verde representa la tasa de petróleo real y la línea naranja la predicción. La diferencia en barriles entre ambos datos se totaliza y se promedia respecto a la cantidad de días de producción. Este procedimiento es realizado por el algoritmo de forma automática.

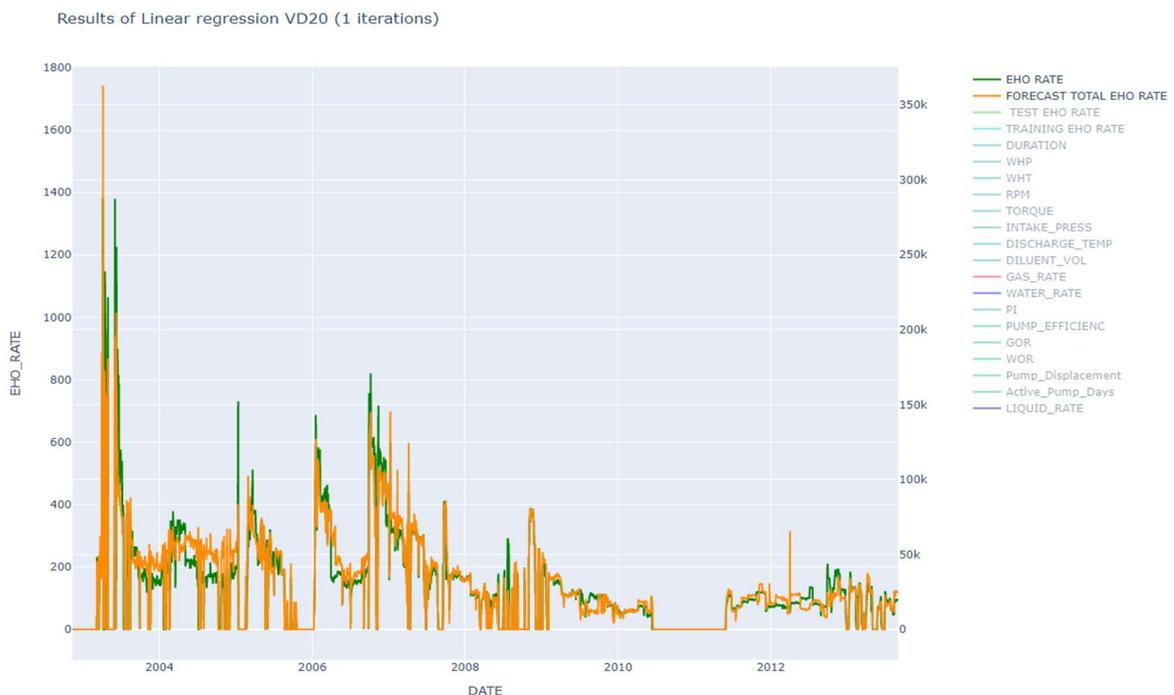


Figura 52: Ejemplo del modelo de predicción de regresión lineal multivariable del pozo VD20

Error promedio de producción sobreestimada y subestimada: determina el promedio de la cantidad de barriles subestimados y sobreestimados, para determinar cuál de estas dos variantes está más presente en el modelo (subestimación o sobreestimación).

Tasa de petróleo total (bpd): totaliza la cantidad de barriles producidos durante todos los días de producción del pozo.

Predicción de la tasa de petróleo (bpd): totaliza la cantidad de barriles calculados en la predicción durante todos los días de producción del pozo.

Error porcentual entre la tasa total real y la predicción: determina el porcentaje de error con la misma ecuación del error relativo, para determinar la relación existente entre ambos parámetros calculados y verificar que tan lejos de la realidad está el modelo.

Adicionalmente a este procedimiento de cálculo de errores, se realizó un proceso para determinar la significancia estadística de cada una de las variables de entrada respecto a la salida calculando el valor de p y así determinar cuáles son los parámetros con mayor influencia en el comportamiento del pozo.

- Regresión polinómica

Esta metodología es similar, aunque un poco más compleja que la regresión lineal multivariable y se explica con detalle en el capítulo II. Los parámetros de entrada, al igual que los casos de estudio para este método, son los mismos que en el método explicado anteriormente, sin embargo, este algoritmo busca ajustar los datos de entrada con base a una ecuación polinómica, el cual realiza un proceso iterativo variando los grados del polinomio. A medida que el grado del polinomio aumenta, aumentan los coeficientes en la ecuación al igual que las variables en la misma, por lo tanto, el algoritmo calcula la relación que tiene cada una de las variables considerando el grado y cómo se relacionan entre sí. Éste, calcula los pesos o coeficientes que multiplican a cada conjunto de variables, para que el resultado de la suma de todos los atributos multiplicados por su respectivo coeficiente se acerque lo más posible a la tasa de petróleo diaria. Este proceso va calculando una nueva ecuación que depende del grado del polinomio, y al finalizar muestra el resultado de todos los grados estudiados por el algoritmo. En la figura 53, se muestra un ejemplo de la gráfica resultante de la regresión polinómica, en color amarillo la predicción de grado 1, el color naranja representa la predicción del polinomio grado 2, en color vinotinto la predicción de grado 3, mientras en color azul se muestra la predicción de grado 4. En esta metodología se esperaba estudiar el comportamiento de los polinomios hasta grado 7, sin embargo, no es posible por la capacidad de cómputo disponible. Cada uno de estos casos fue analizado por

separado con la finalidad de definir cuál es el polinomio con mejor ajuste. Mejor colocar el eje y en logarítmico para poder ver algo. Grado 3 y 4 comete picos de errores. Lo otro es hacer grado uno y dos y en otra figura grados 3 y 4.

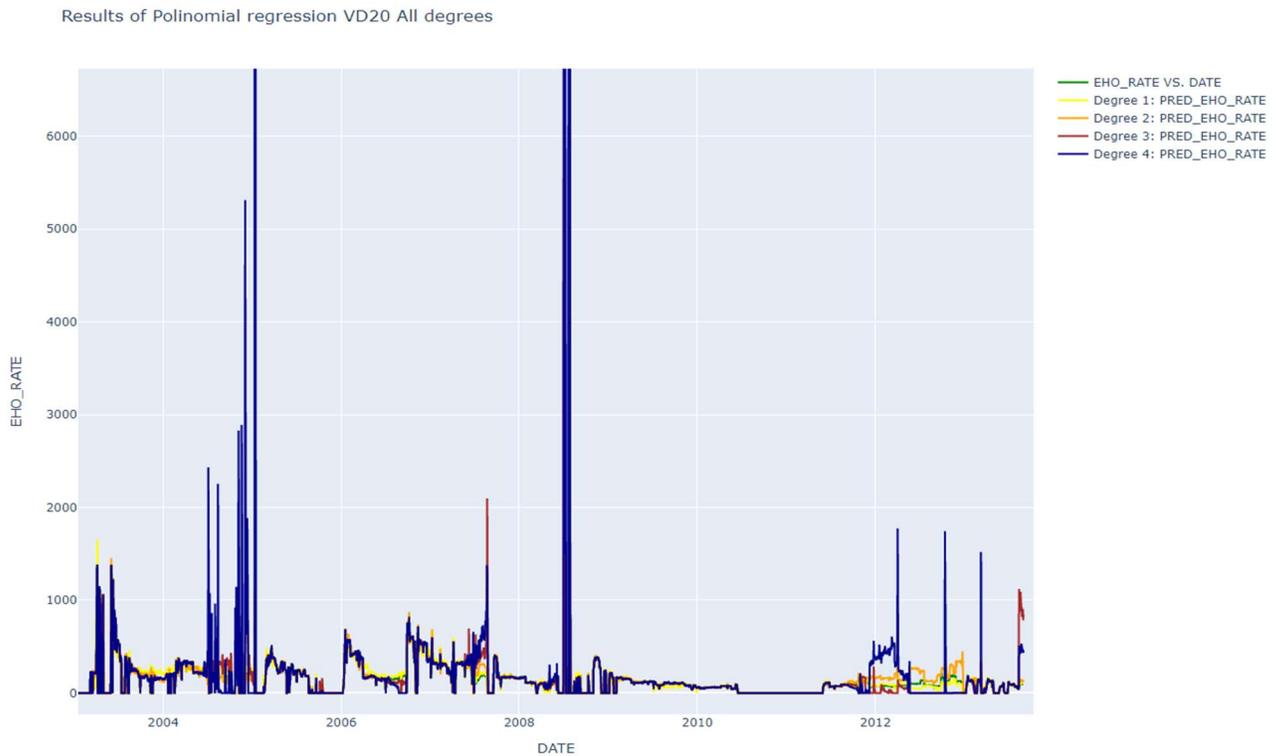


Figura 53: Ejemplo del resultado de la predicción del pozo VD20 por regresión polinómica comparando los grados del polinomio.

- Redes neuronales

El modelo de redes neuronales es el más complejo, por lo tanto, su entrenamiento requiere de diferentes parámetros de funcionamiento, tales como, funciones internas, arquitectura, formato de datos de entrada, entre otros. Las redes neuronales se basan en un modelo iterativo en el cual fue necesario variar todos los parámetros hasta encontrar la combinación óptima que se ajustara a los resultados esperados. Entre los parámetros variados se encuentran: los datos de entrada, escala de datos, número de capas de la red

neuronal, neuronas por capa, función de pérdida, función de activación, optimizador y epochs.

- Datos de entrada: fué uno de los parámetros fundamentales en el diseño y aplicación de redes neuronales, ya que, dependiendo de la calidad y formato de los mismos, el modelo puede obtener un mejor desempeño y lograr obtener resultados en menor tiempo. Los datos de entrada en este método fueron los mismos seis (06) *datasets* usados en los modelos anteriores.
- Escala de datos: Al momento de usar datos reales, los valores más elevados como la tasa de gas tendrán mayor influencia sobre el parámetro de salida que aquellos parámetros más pequeños como la eficiencia de la bomba. Por esta razón se trabaja de igual forma con los *datasets* estandarizados y normalizados, con la finalidad de comparar y verificar si los cálculos que se realizaron en la red neuronal son más rápidos y la influencia que cada parámetro de entrada tiene sobre el resultado es equitativo.
- Número de capas de la red neuronal: El número de capas intermedias fueron muy importantes en el modelo, ya que fueron la base estructural de las neuronas encargadas de realizar los cálculos desde la capa de entrada hasta la capa de salida. En este TEG se decidió variar el número de capas intermedias desde una (01) capa a tres (03) capas, ya que, es el número de capas común en tareas simples de clasificación o regresión, colocar más capas intermedias puede generar un sobre entrenamiento del modelo, por lo general modelos con más capas son usados para procesamiento de imágenes y videos, lo cual no es parte de esta investigación.
- Neuronas por capa: es otro de los parámetros básicos que se varió en el modelo de RNA. El número de neuronas de entrada de la red dependió directamente del número de variables de entrenamiento, sin embargo, este número aumenta para cada pozo de acuerdo con la cantidad de cambios de bomba que se realicen durante el periodo de producción del mismo. Los parámetros a predecir con este método fueron: la producción de agua (*WATER_RATE*) y la producción de petróleo (*EHO_RATE*), por lo tanto, se establecieron dos (02) neuronas de salida. El número de neuronas en las capas ocultas es definido por el usuario, sin embargo, no existe

una forma directa para definir cuantas neuronas de este tipo son necesarias para un trabajo determinado, y, por lo tanto, es considerado un trabajo de ensayo y error. Para este procedimiento se realizaron varias pruebas y se fue ajustando el número de neuronas en la capa oculta de acuerdo con los resultados obtenidos. Al igual que con el número de capas, la cantidad de neuronas en cada capa no debe ser muy elevado, ya que el diseño será muy robusto y el modelo resultaría sobreentrenado. Si bien no hay una forma directa de saber el número de neuronas óptimo para una RNA varios autores han desarrollado algunas reglas que permiten tener una idea aproximada del número de neuronas que se debería tener para obtener resultados aceptables, sin embargo, éstas no son determinantes para un modelo de RNA y pueden variar de acuerdo a cada esquema. En esta investigación se usaron 6 valores para estimar ese número de neuronas, cuatro valores calculados (04) que dependen del número de neuronas de entrada y de salida, el número de muestras en el set de entrenamiento y un factor de escala arbitrario que permita ajustarse lo más posible a los resultados que se quiere, mientras los otros dos valores fueron fijos entre el rango de estudio los cuales son cincuenta ($N_{h1} = 50$) y doscientos ($N_{h2} = 200$). Las ecuaciones utilizadas para determinar el número de capas a utilizar en el modelo son las siguientes:

$$\circ N_{h3} = \frac{N_s}{(\alpha * (N_i + N_o))} \quad Ec. 12$$

$$\circ N_{h4} = \frac{N_i + N_o}{2} \quad Ec. 13$$

$$\circ N_{h5} = \left(\frac{2}{3} * N_i\right) + N_o \quad Ec. 14$$

$$\circ N_{h6} = (2 * N_i) - 1 \quad Ec. 15$$

Donde:

N_i = Número de neuronas de entrada

N_o = Número de neuronas de salida

N_s = Número de muestras en el dataset de entrenamiento

α = factor de escala arbitrario, el valor usado fue $\alpha = 3$.

- Función de pérdida o Loss Function: Dado que la RNA va a ser usada como método de regresión, se espera que los valores de salida sean números reales, por lo tanto, se analiza el desempeño de la red neuronal usando las dos funciones de costo definidas en el capítulo II: *Mean Squared Error* y *Mean Absolute Error*. Ambas funciones representan el error entre los valores predichos y los valores reales, por lo tanto, la red busca minimizar ambos errores lo más posible.
- Función de activación: En este estudio, la función de activación seleccionada fue ReLU (*Rectified Linear Unit*), dado que es un problema de regresión, cuyos valores de salida tienen que ser números reales y mayores o iguales a cero.
- Optimizador: para este modelo se usó el optimizador ADAM (*Adaptive Moment Estimation*), ya que es el más usado por su versatilidad y eficiencia. Además, este optimizador no requiere mucho poder computacional ni mucha memoria durante su implementación.
- Epochs: otro aspecto clave de una RNA es la cantidad de epochs o iteraciones en que la red se entrena. Si el modelo se entrena en pocas iteraciones, este podría no ajustarse adecuadamente, mientras que, si se realiza en muchas iteraciones, podría ocurrir un sobreajuste (*overfitting*). Al igual que con los parámetros anteriores, no existe una cantidad predeterminada de iteraciones para un problema dado, también es considerado un proceso de ensayo y error. En problemas con grandes cantidades de datos normalmente se requieren miles de epochs para un buen entrenamiento, mientras que, en este caso, como se va a aplicar una RNA a cada pozo estudiado y la cantidad de datos analizados en cada pozo no es muy grande, se decidió establecer una cantidad de 1500 iteraciones, y dependiendo de los resultados obtenidos con esta cantidad, se evaluarían otras opciones (más o menos epochs,

dependiendo del caso). Con la finalidad de ajustar la cantidad de iteraciones necesarias, se requirió analizar las curvas de aprendizaje. En ellas se observa mejor si un modelo esta sobreenentrenado o si se podría minimizar aún más el error modificando el número de iteraciones.

En la figura 54 se muestra la estructura base de la RNA a usar en esta metodología, la cual comprende del lado izquierdo la capa de entrada y en su interior las neuronas de entrada con los 14 parámetros base, en la parte central se encuentran las capas ocultas, que en este caso se representan tres (03) capas con sus respectivas neuronas intermedias y del lado derecho se encuentran la capa de salida con las neuronas de salida, mientras cada una de las líneas que las conecta representa el peso de cada conexión entre neuronas.

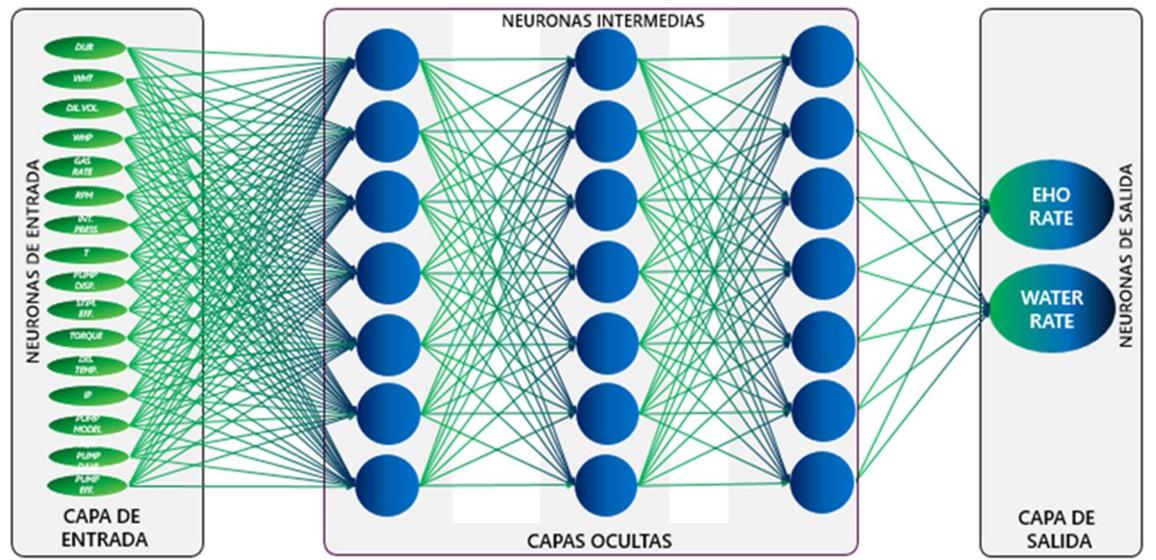


Figura 54: Estructura base de la RNA con tres capas ocultas

Una vez que todos los parámetros fueron establecidos, se procedió a modelar las RNA e ir evaluando los mismos. Es necesario tomar en cuenta todas las combinaciones posibles de acuerdo con los parámetros, por lo tanto, se requiere modelar la siguiente cantidad de RNA:

$$N_{RNA/Pozo} = N_{Datos} * N_{F.P.} * N_{Neuronas} * N_{Capas} * N_{Escala} \quad \text{Ec. 16}$$

Donde:

$N_{\text{RNA/Pozo}}$ = Numero de modelos de redes neuronales por cada pozo
 N_{Datos} = Cantidad de Sets de datos distintos (2) – Original, caracterizado
 $N_{\text{F.P.}}$ = Cantidad de funciones de pérdida distintas (2) – MSE, MAE
 N_{Neuronas} = Sets de neuronas para las capas ocultas (6)
 N_{Capas} = Numero de distintas configuraciones para el número de capas ocultas (3) – 1, 2, 3
 N_{Escala} = Número de distintos tipos de manipulación de datos (3) – Normal, estandarizado, normalizado

Basados en este cálculo fue necesario realizar 216 modelos de RNA para cada uno de los pozos de la muestra, y así determinar la combinación de parámetros óptima para esta metodología, generando así un total de 8.640 modelos de RNA. Desafortunadamente una de las limitaciones de este TEG es la capacidad de cómputo y a pesar de que se trabaja con servidores externos, la conexión a internet no es estable por largos periodos de tiempo, por lo tanto, se hizo imposible llevar a cabo esta estrategia, y se procedió a diseñar una nueva estrategia que permitió reducir la cantidad de modelos necesarios y obtener los mejores resultados posibles.

Este nuevo método se basa en realizar varios modelos de RNA para un mismo pozo dejando fija la cantidad de capas ocultas, las neuronas en cada capa oculta y el número de epochs, por lo tanto, se definen 18 tipos de RNA, los cuales se detallan a continuación en la tabla 10.

Tabla 10: Esquema de combinaciones RNA

Típos	Capas ocultas	Numero de Neurona
RNA 1	1	$N_{h1} = 50$
RNA 2	2	$N_{h1} = 50$
RNA 3	3	$N_{h1} = 50$
RNA 4	1	$N_{h2} = 200$
RNA 5	2	$N_{h2} = 200$
RNA 6	3	$N_{h2} = 200$
RNA 7	1	N_{h3}
RNA 8	2	N_{h3}
RNA 9	3	N_{h3}
RNA 10	1	N_{h4}
RNA 11	2	N_{h4}
RNA 12	3	N_{h4}
RNA 13	1	N_{h5}
RNA 14	2	N_{h5}
RNA 15	3	N_{h5}
RNA 16	1	N_{h6}
RNA 17	2	N_{h6}
RNA 18	3	N_{h6}

Una vez definidos los 18 tipos de RNA, se seleccionaron 10 pozos de manera aleatoria los cuales corresponden al 25% de la muestra, 5 del yacimiento tipo deltaico y 5 del yacimiento tipo fluvial. Basados en esta selección se planteó la metodología para definir la configuración óptima de los parámetros en cada pozo. En la primera etapa se toman los dos *datasets* con datos reales, tanto originales como caracterizados para estudiar su respuesta con las dos funciones de pérdida (MAE y MSE), resultando cuatro casos para cada pozo con los 18 tipos de RNA, los cuales se detallan a continuación en la tabla 11.

Tabla 11: Configuración de parámetros para la etapa 1

Casos	Set de datos	Función de pérdida	Cantidad de RNA	Errores
1	Reales originales	MAE	18	Absoluto / Relativo
2	Reales originales	MSE	18	Absoluto / Relativo
3	Reales caracterizados	MAE	18	Absoluto / Relativo
4	Reales caracterizados	MSE	18	Absoluto / Relativo

Con base en los 18 resultados que se obtienen en cada uno de los casos antes planteados, se seleccionan los 5 modelos de RNA con menor error relativo promedio y menor error absoluto en cada caso. Sin embargo, es necesario estudiar la eficiencia de cada uno de los casos respecto a su set de datos y a la función de pérdida, ya que en cada uno de los pozos el modelo de RNA seleccionado puede ser diferente y por lo tanto no comparable. Por esta razón, se calcula el error relativo promedio y absoluto de cada caso, tomando en cuenta los 5 mejores modelos. En la figura 55a se observa la selección de los 5 modelos con menor error absoluto promedio entre los datos reales y la predicción del modelo, mientras en la figura 55b se observa un ejemplo de la selección de los 5 mejores modelos con menor error relativo promedio, como se observa, no todos los modelos cumplen con tener un bajo error relativo y un bajo error absoluto, por lo tanto se tomaron en cuenta los modelos que cumplan con ambos criterios como se observa en la figura 55c, y con este resultado se determina el error promedio de cada caso, con la finalidad de comparar los resultados obtenidos y determinar entre los cuatro cual es la función de pérdida y set de datos con menor error.

(a) Selección de menor error absoluto promedio							(b) Selección de menor error relativo promedio						
Loss func	Neurons/Layer	Layers	Epochs	time	mae	mape	Loss func	Neurons/Layer	Layers	Epochs	time	mae	mape
mae	7	1	1500	115.41	88.29	13.89	mae	7	1	1500	115.41	88.29	13.89
mae	7	3	1500	130.53	120.90	22.25	mae	7	3	1500	130.53	120.90	22.25
mae	7	2	1500	121.61	182.25	40.17	mae	7	2	1500	121.61	182.25	40.17
mae	10	3	1500	124.51	113.81	20.22	mae	10	3	1500	124.51	113.81	20.22
mae	10	2	1500	120.14	109.42	23.51	mae	10	2	1500	120.14	109.42	23.51
mae	10	1	1500	114.75	147.85	28.35	mae	10	1	1500	114.75	147.85	28.35
mae	27	3	1500	127.59	165.13	25.88	mae	27	3	1500	127.59	165.13	25.88
mae	27	1	1500	113.85	150.38	26.89	mae	27	1	1500	113.85	150.38	26.89
mae	27	2	1500	125.03	230.50	47.48	mae	27	2	1500	125.03	230.50	47.48
mae	50	3	1500	132.79	175.06	33.30	mae	50	3	1500	132.79	175.06	33.30
mae	50	2	1500	124.39	185.27	35.45	mae	50	2	1500	124.39	185.27	35.45
mae	50	1	1500	118.12	253.64	48.28	mae	50	1	1500	118.12	253.64	48.28
mae	61	2	1500	128.84	180.30	28.55	mae	61	2	1500	128.84	180.30	28.55
mae	61	3	1500	137.69	186.73	38.39	mae	61	3	1500	137.69	186.73	38.39
mae	61	1	1500	114.06	218.60	39.52	mae	61	1	1500	114.06	218.60	39.52
mae	200	2	1500	164.15	167.48	25.12	mae	200	2	1500	164.15	167.48	25.12
mae	200	3	1500	207.55	178.85	34.66	mae	200	3	1500	207.55	178.85	34.66
mae	200	1	1500	120.14	250.78	45.93	mae	200	1	1500	120.14	250.78	45.93

(c) Mejores casos de RNA en el pozo de estudio con menor error relativo y promedio para una función de pérdida dada.

Loss func	Neurons/Layer	Layers	Epochs	time	mae	mape
mae	7	1	1500	115.41	88.29	13.89
mae	7	3	1500	130.53	120.90	22.25
mae	10	3	1500	124.51	113.81	20.22
mae	10	2	1500	120.14	109.42	23.51

Figura 55: Selección de las mejores RNA en el pozo VD20

Una vez seleccionada la configuración de parámetros óptima de los cuatro (04) primeros casos, se procedió a la segunda etapa donde se realizó el mismo procedimiento que en la primera etapa, pero variando en esta, la escala de los datos, generando tres (03) nuevos casos, como se observa en la tabla 12, la cual muestra el resultado del caso base, (estos resultados pueden variar de acuerdo con cada pozo).

Tabla 12: Ejemplo de configuración de parámetros para la etapa 2

Casos	Set de datos	Función de pérdida	Cantidad de RNA	Errores
1	Reales caracterizados	MAE	18	Absoluto / Relativo
2	Caracterizados normalizados	MAE	18	Absoluto / Relativo
3	Caracterizados estandarizados	MAE	18	Absoluto / Relativo

Al finalizar la etapa segunda etapa, se obtiene la configuración de parámetros óptima de un pozo. Este procedimiento se repitió para los 10 pozos seleccionados con la finalidad de definir si la configuración óptima de los parámetros encontrada es

recurrente, y puede ser aplicada al resto de la muestra, en caso de no serlo, deberá aplicarse esta metodología para todos los pozos en estudio. En la figura 56, se observa un esquema base de las etapas de este método, donde los cuadros de color azul representan los cuatro primeros casos explicados cada uno con 18 tipos de RNA asociadas, para un total de 72 RNA en la primera etapa. El cuadro de salida en color naranja representa la configuración de parámetros óptima de los cuatro primeros casos y esta configuración de salida es tomada como referencia para la segunda etapa. En esta nueva etapa se corrieron dos nuevos casos con la variación de escala, los cuales se representan en color azul, generando 36 nuevos tipos de RNA. Por último, de la comparación de estos tres últimos casos, se obtiene la configuración óptima de parámetros, disminuyendo la cantidad de RNA a 108 por cada pozo.

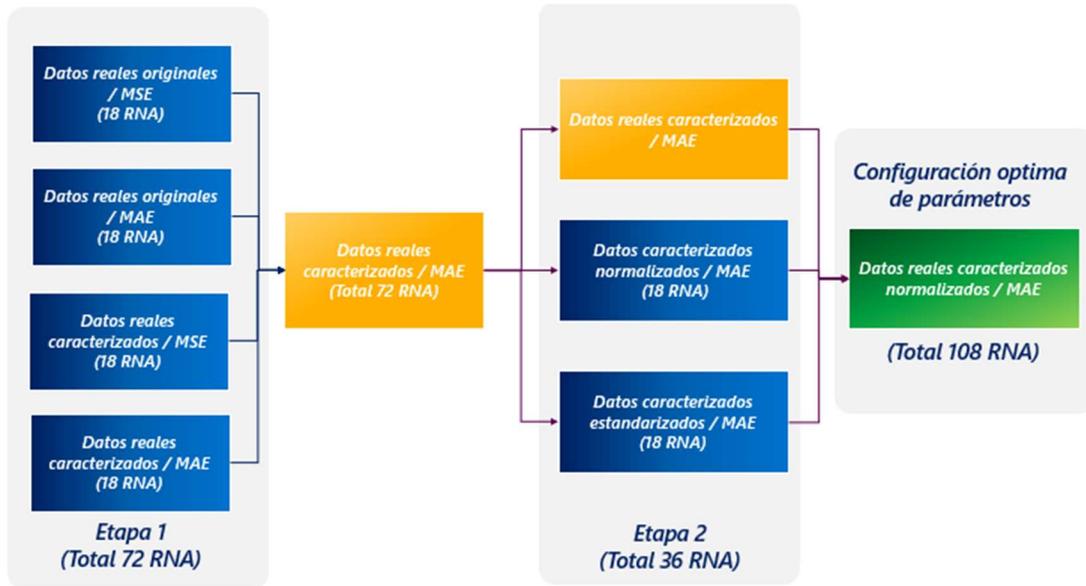


Figura 56: Esquema de la metodología para definir la configuración óptima de parámetros de una RNA.

Al finalizar este procedimiento, se volvieron a procesar los 18 tipos de RNA para el 100% de la muestra, pero esta vez, solo con la configuración de parámetros seleccionada, con la finalidad de hallar un único modelo para cada pozo, el cual tenga el menor error absoluto y relativo respecto a su predicción. Este modelo se guardó, incluyendo su arquitectura, pesos, funciones internas, entre otros, y se le asignó una etiqueta característica para cada pozo que puede ser utilizado en cualquier momento.

Los tres modelos antes expuestos (lineal, polinomial y redes neuronales) fueron comparados con la finalidad de obtener un único modelo con el mejor ajuste para cada uno de los pozos y así poder obtener una predicción de la producción óptima. Es importante tomar en cuenta que en cada uno de los modelos se obtiene una ecuación característica, la cual relaciona el comportamiento de las BCP con las otras características del pozo. Por lo tanto, a través de estas similitudes se construyó una nueva base de datos donde se almacena información mensual de la producción de petróleo, gas y agua, en conjunto con la información de la bomba activa, como el desplazamiento nominal y su coeficiente característico. Esta nueva tabla permitió generar un nuevo modelo de predicción de tasa para nuevas bombas que no hayan sido probadas en el pozo.

El modelo de K-nearest neighbors buscó agrupar y clasificar los pozos de acuerdo a sus características de producción, por lo tanto, al momento de tratar de predecir el comportamiento con una bomba no utilizada en el pozo, el algoritmo tiene la capacidad de buscar el coeficiente de la bomba en el grupo de pozos con las mayores similitudes de producción y en caso de que exista algún entrenamiento previo de la bomba en otro pozo, el algoritmo tendrá la capacidad de predecir la producción en el pozo de estudio basado en la experiencia de otro pozo.

Al finalizar el proceso de modelado de predicción, se procedió a realizar el modelo de análisis de las BCP.

3.2.4.2.- Modelado del comportamiento de las BCP

Para el modelado del comportamiento de las BCP se utilizaron 3 metodologías, KNN, *support vector machine* y árbol de decisiones que se explican en el capítulo II, con la finalidad de determinar el estado de funcionamiento de la bomba.

Este tipo de metodologías son consideradas parte del aprendizaje supervisado, ya que se basó en el estudio de anomalías, las cuales deben ser especificadas y estudiadas por el usuario bajo su experiencia profesional. El algoritmo final logró identificar las fallas y una vez que estas fueron etiquetadas, el mismo, tiene la capacidad de determinar fallas de este tipo en el futuro.

El primer paso de esta metodología fue estudiar cada una de las bombas como entes independientes, por lo tanto, el historial de producción de los pozos se dividió de acuerdo con la cantidad de bombas del mismo. Por otra parte, la programación base de todas las metodologías se basó en los parámetros principales de la bomba, como lo son el torque, RPM, cantidad de días activos de la bomba, las tasas de producción y el desplazamiento nominal de la bomba, el cual es diferente para cada modelo.

En la figura 57, se observa la tabla base de las bombas correspondientes al pozo tipo (VD20), donde se muestra el modelo de la bomba, el número de bomba al cual corresponde en el pozo (etiqueta única), fecha de inicio y final de producción, *p-value*, cantidad de días activos, desplazamiento nominal de cada bomba en bpd/RPM, cantidad total de petróleo producido, petróleo producido promedio por RPM y promedio de petróleo producido por día. Estos datos permiten dar un pequeño resumen de la eficiencia de cada bomba, sin embargo, el entrenamiento es más complejo.

	Pump_Model	Pump No.	Start Date	End Date	P values	Active Pump Days	Pump Displacement	EHO RATE produced	EHO_RATE / RPM	EHO RATE p/d
0	430TP2000	Pump 1	2002-11-12	2005-02-12	0.0000	540.0	5.1575	141823.984260	4.151431	262.637008
1	24.40-1200	Pump 2	2005-02-13	2005-12-29	0.0000	185.0	2.4000	46983.709932	1.179811	253.966000
2	20.40-2100	Pump 3	2005-12-30	2006-09-25	0.0078	258.0	4.2000	62371.350984	0.911607	241.749422
3	580TP1600	Pump 4	2006-09-26	2007-09-03	0.0000	313.0	6.8875	104108.447700	1.345236	332.614849
4	22.40-2500	Pump 5	2007-09-04	2008-08-07	0.0000	274.0	5.0000	39838.765723	6.246502	145.396955
5	20.40-2100	Pump 6	2008-08-08	2013-01-04	0.0078	1126.0	4.2000	117119.110343	1.061105	104.013419
6	430TP2000	Pump 7	2013-01-05	2013-02-18	0.0000	38.0	5.1575	4648.368487	0.473911	122.325486
7	750TP1200	Pump 8	2013-02-19	2013-03-16	0.0001	18.0	8.9975	2033.964987	0.466354	112.998055
8	430TP2000	Pump 9	2013-03-17	2013-05-19	0.0000	36.0	5.1575	3625.194296	0.479788	100.699842
9	-	Pump 10	2013-05-20	2013-06-24	0.6959	20.0	0.0000	1464.973674	0.511379	73.248684
10	28.40-500	Pump 11	2013-06-25	2013-08-31	0.0021	65.0	1.0700	5421.109352	0.690554	83.401682

Figura 57: Tabla de descripción de las bombas en el pozo VD20

Para el entrenamiento se seleccionan los mismos periodos de entrenamiento utilizados para el modelado de predicción, con la finalidad de buscar patrones entre las variables que permitan determinar fallas en el sistema. Una variable fundamental corresponde a la velocidad en RPM, por lo tanto, se busca determinar la tasa de producción de líquido ideal diariamente, multiplicando la tasa nominal de la bomba por las RPM dadas y luego este se compara de forma porcentual con la tasa de petróleo. Por otro lado, se busca determinar por medio de regresiones la relación que existe entre la tasa de petróleo real y la influencia de otros parámetros respecto a la tasa ideal. El promedio porcentual de estas diferencias y relaciones entre las regresiones indican la eficiencia porcentual de la bomba a la profundidad dada respecto a su tasa ideal.

En la figura 58, la línea de color azul representa la tasa ideal a condiciones de superficie, la línea verde representa la tasa real de petróleo, mientras la línea de color naranja representa la tasa ideal a condiciones de subsuelo y la línea en color vinotinto representa las RPM diarias. Por otro lado, se observa una línea horizontal punteada color morada, la cual es una guía para identificar que en ambos periodos (periodo 1 y periodo 2), las RPM se mantienen constantes, sin embargo, se observa que en el periodo 1 la tasa real está por debajo de la ideal, mientras que en el periodo 2 el caso es el contrario. Por esta razón fue necesario calcular una tasa ideal a condiciones reales que sea solo dependiente de las RPM y así determinar casos de anomalías respecto a otros parámetros como: torque, temperaturas, tasa de agua o gas, entre otros.

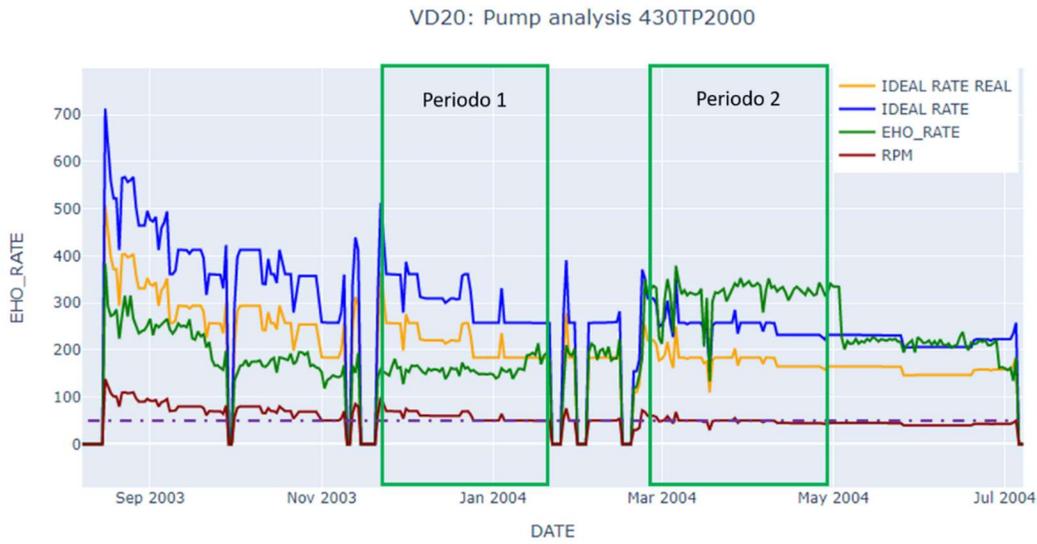
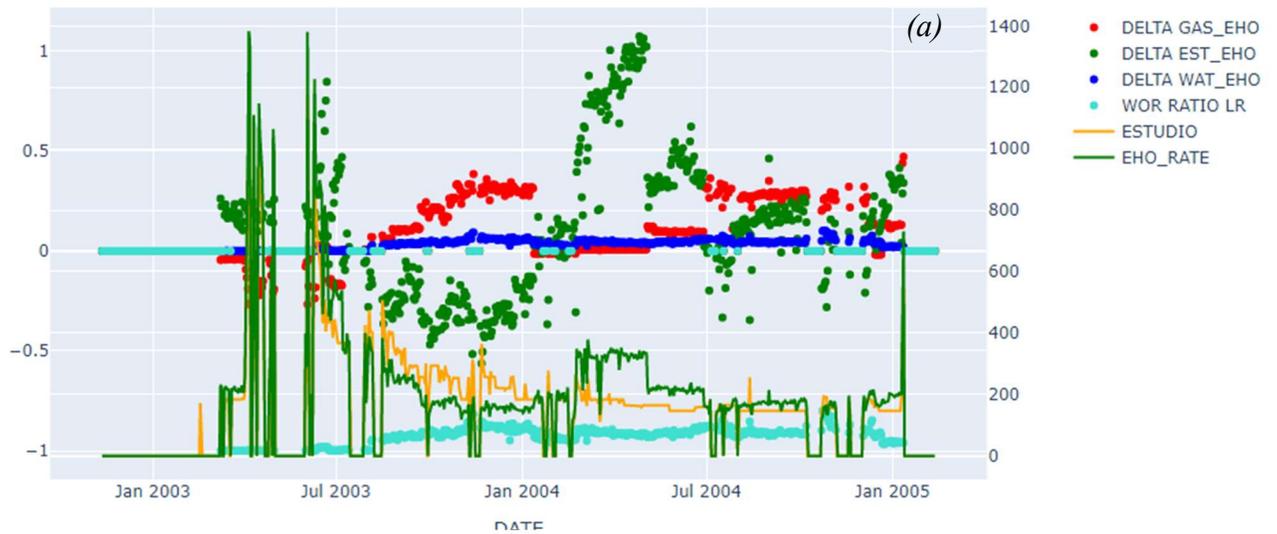


Figura 58: Calculo de tasa ideal en la bomba 1 del pozo VD20

A través de la normalización de las tasas de producción de gas, petróleo y agua, al igual que la tasa ideal antes expuesta, se estudió la relación de cambio de la tasa de petróleo respecto a cada una, con la finalidad de encontrar patrones, que al ser comparados con el comportamiento ideal demuestren algún tipo de anomalía.

En la figura 59a, se observa la gráfica resultante del estudio del comportamiento de la primera bomba del pozo tipo, donde el eje “x” muestra la fecha de estudio, el eje “y” principal muestra la escala de los parámetros normalizados, los cuales se muestran con puntos a lo largo del estudio y el eje y secundario muestra la escala real de la tasa de producción los cuales se representan con líneas continuas.

VD20: Pump analysis 430TP2000



VD20: Pump analysis 430TP2000

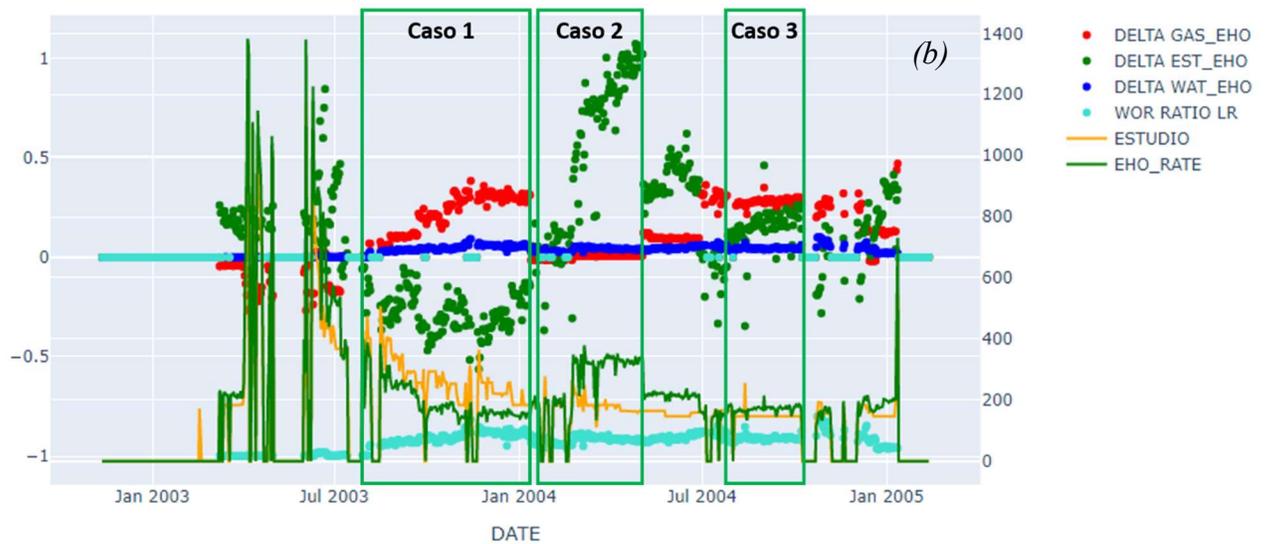


Figura 59: (a) Gráfica resultante del estudio del comportamiento bomba 1 del pozo tipo (b) Estudio de casos de acuerdo con el análisis de parámetros

El primer parámetro representado con puntos en color rojo (“DELTA GAS_EHO”) muestra la diferencia existente entre la tasa de gas y la tasa de petróleo normalizado.

$$.Delta Gas_Eho = Gas_rate_{norm} - Eho_rate_{norm}. \quad Ec. 17$$

El segundo parámetro representado con puntos de color verde (“DELTA EST_EHO”) muestra la diferencia entre la tasa de petróleo real y la tasa ideal de producción de la bomba a condiciones de subsuelo respecto a la tasa ideal

$$Delta\ est_eho = \left(\frac{Eho_rate - Idea_rate}{Ideal_rate} \right) \quad Ec. 18$$

El tercer parámetro representado con puntos en color azul oscuro (“DELTA WAT_EHO”) muestra la relación de producción de agua respecto a la tasa de líquido.

$$Delta\ wat_eho = \frac{Water_rate}{Liquid_rate} \quad Ec. 19$$

Mientras el cuarto parámetro representado en color turquesa (“WOR RATIO LR”) muestra la diferencia agua petróleo respecto a la tasa de líquido total

$$WOR\ Ratio\ LR = \frac{Water_rate - Eho_rate}{Liquid_rate} \quad Ec. 20$$

El quinto parámetro en color amarillo y línea continua muestra la tasa ideal de la bomba a condiciones de superficie y la línea verde muestra la tasa real.

Todos estos parámetros permitieron dividir la historia de producción en distintos casos y los algoritmos programados debían identificar cada caso de acuerdo a los datos suministrado. La figura 59b muestra un ejemplo donde existen 3 casos claros: el primer caso corresponde a un caso con alta producción de gas, menos producción de petróleo y menor rendimiento de la bomba respecto a la tasa ideal, el caso 2 representa un buen rendimiento de la bomba respecto a la tasa ideal con una ¿ relación gas petróleo, mientras el caso 3 muestra una caída de producción al límite ideal que podría estar ocasionado por un aumento de la tasa de gas. En los tres casos se puede observar que el comportamiento de agua no tiene un efecto importante sobre la producción de petróleo para este período.

Este tipo de análisis se realizó con cada una de las bombas de la muestra aplicando las 3 metodologías.

Para mostrar los resultados de este estudio, se programó una gráfica interactiva basada en una escala de color que muestra el estado de la bomba y algunos comentarios importantes, según el caso.

Con estos casos definidos y el estado de la bomba se procedió a visualizar la tabla de significancia estadística que se obtiene en el modelado de predicción, con la finalidad de evaluar los parámetros con mayor significancia estadística y aquellos que pudiesen tener un efecto o impacto sobre el comportamiento de la bomba y así encontrar nuevas anomalías o la causa del problema.

Todos estos parámetros y el estudio detrás de cada bomba son muy importantes para un correcto aprendizaje del algoritmo y un análisis del funcionamiento de las BCP.

A continuación, se explica la última fase de la metodología CRISP-DM correspondiente al despliegue que comprende la creación de una interfaz gráfica para cumplir con el objetivo general de esta investigación. Posteriormente, en el capítulo IV, se muestran los resultados obtenidos de ambos modelos y la evaluación de los mismos.

3.2.5.- Despliegue

En la última fase de despliegue se generó una interfaz gráfica donde se le permite al usuario realizar cada uno de los procedimientos aquí explicados de forma rápida y sencilla. Esta interfaz fue creada con una biblioteca gráfica para el lenguaje de programación de Python, en la cual se combinan todos los algoritmos de análisis desarrollados a lo largo de este TEG.

En el siguiente capítulo se muestra el análisis de los resultados obtenidos en cada una de las fases y etapas antes expuestas.

CAPÍTULO IV

ANÁLISIS DE RESULTADOS

En este capítulo se exponen los resultados de la metodología presentada en el capítulo III. Con la finalidad de cumplir con los objetivos propuestos este análisis se realizará en cuatro partes: el análisis de la selección de las metodologías de *machine learning* y *data science*, el análisis de los modelos de predicción, la significancia estadística y el análisis del modelado del comportamiento de las BCP.

4.1.- Análisis de la selección de las metodologías de *machine learning* y *data science*

El primer objetivo de esta investigación se basó en revisar toda la información bibliográfica acerca de herramientas de análisis de datos como, *Data Science* y *Machine Learning* y su aplicación en la industria petrolera. En Venezuela son pocas las empresas en la industria petrolera que han aplicado este tipo de metodologías. Sin embargo, a nivel internacional, varias compañías como Schlumberger, Devon Energy, Chevron y Equinor, entre otras, han buscado mejorar sus operaciones con el uso de la ciencia de datos y el aprendizaje automatizado. Estas compañías han logrado reducir el tiempo de inoperatividad en pozos, incrementar la producción en pozos horizontales, disminuir los costos operativos, realizar un modelo predictivo de completación de pozos, predecir la producción de yacimientos no convencionales, entre otros.

Como resultado de la investigación realizada se puede decir que no existe una herramienta definida para la aplicación de este tipo de metodologías y que dependen del objetivo final del proyecto. Sin embargo, se encontraron dos metodologías de ciencia de datos que pueden ser aplicadas a cualquier área, SEMMA y CRISP-DM. Durante el proceso de análisis de ambas metodologías se determinó que la metodología más completa y que se ajustaba a los objetivos de esta investigación era la metodología CRISP-DM, ya que, comprendía todas las etapas de desarrollo requeridas, desde la interpretación del problema, el análisis de los datos, el desarrollo del modelado y el

despliegue, que garantiza que la investigación y el desarrollo de las herramientas sean de fácil uso para la empresa o cualquier usuario.

Durante el análisis de los datos de producción y operatividad, al igual que durante el estudio de las herramientas de *machine learning*, se pudo determinar que no existe una única herramienta de aprendizaje automatizado que pudiese generar resultados satisfactorios para la investigación. Por esta razón, se realizó un conglomerado de diferentes herramientas. Para los modelos de predicción se utilizaron herramientas de regresión que fuesen capaces de devolver valores numéricos mientras que para los modelos de análisis de comportamiento de las BCP fue necesario usar otro tipo de herramientas que fuesen capaces de clasificar en base a una serie de parámetros. Las herramientas de regresión seleccionadas fueron tres: regresión lineal multivariable, regresión polinomial y redes neuronales. Por otra parte, las herramientas de clasificación seleccionadas fueron: *support vector machine*, KNN y árbol de decisiones. Todas estas metodologías se estudiaron y analizaron de manera individual para definir cuál era la que mejor se ajustaba a los objetivos y se explicaran más adelante en este capítulo.

La macolla VD seleccionada posee 47 pozos, sin embargo, el historial de producción recibido contiene pocos datos de los últimos 7 pozos perforados, por lo tanto, no se pudo aplicar ningún modelo de predicción a estos pozos.

La falta de organización y estandarización en las bases de datos entregadas fue uno de los principales problemas- Otro de los problemas encontrados fue la escasez de datos respecto a algunos pozos y parámetros que no permitieron un buen análisis de datos.

El desarrollo de los algoritmos requiere que la información esté estructurada, por lo tanto, es imposible trabajar este tipo de metodologías con las bases de datos originales. Por otro lado, los vacíos de información o la falta de datos aumentan la incertidumbre y disminuyen la precisión de los modelos generados. Los resultados antes mencionados se explican a profundidad en las siguientes secciones de este capítulo.

4.2.- Análisis de los modelos de predicción

La primera parte de este TEG se basó en obtener el modelo de predicción adecuado que describiera el comportamiento de producción de cada pozo. Se estudiaron los tres casos antes mencionados, variando los *datasets* con dos características principales. Con la primera característica se estudió cómo afecta la calidad del dato al modelo (datos originales y datos caracterizados), mientras que con la segunda característica se estudió cómo afecta el escalamiento en el modelo y en el tiempo de cómputo.

Estas tres metodologías se estudiaron y analizaron de manera individual, para luego definir cuál es la herramienta que mejor se ajusta a los objetivos y obtuvo el menor error.

4.2.1.- Regresión lineal multivariable

El modelo de predicción de regresión lineal multivariable fue el modelo más sencillo de los tres en estudio. El resultado de este modelo devolvió una ecuación lineal característica para cada pozo, con un promedio de 19 coeficientes por pozo, el cual fue proporcional a la cantidad de parámetros de entrada.

El algoritmo procesó en promedio un total de 48416 datos por pozo para un ajuste cuadrático promedio de $r^2 = 0.9266$.

La muestra seleccionada (macolla VD) contenía un total de 40 pozos válidos distribuidos en cuatro arenas, por lo tanto, para simplificar los resultados del modelo en cada pozo, estos se agruparon y se promediaron de acuerdo con la arena a la que pertenecían y en base a las dos características en estudio, obteniendo los siguientes resultados mostrados en la tabla 13.

Tabla 13: Resultados (error) de la selección del set de datos para regresión lineal multivariable

Yacimiento / Set de datos	Reales		Normalizados		Estandarizados	
	Original	Caracterizado	Original	Caracterizado	Original	Caracterizado
B2	23.87%	23.62%	33.77%	32.12%	33.74%	23.84%
C2	17.96%	15.29%	31.93%	31.05%	37.01%	34.43%
D3	30.91%	27.28%	36.84%	34.46%	44.78%	35.96%
E1	20.36%	15.41%	43.77%	23.65%	31.93%	35.93%
Promedio	23.28%	20.40%	36.58%	30.32%	36.87%	32.54%

En los resultados de la tabla 13, se observa que, respecto a la primera característica de los sets de datos, los datos caracterizados tienen un menor error en la mayoría de los casos que los datos originales.

Para el set de datos reales, el set original tiene un error promedio de 23.28 %, mientras el set caracterizado tiene un error de 20.4%, con una diferencia de casi 3% para este conjunto.

El set de datos normalizados refleja la misma tendencia, con los datos originales normalizados el error relativo promedio es de un 36.58 %, mientras que para los datos caracterizados normalizados el error disminuye a un 30.32%, con una diferencia del 6%.

Por último, para el set de datos estandarizados, los originales obtuvieron un error relativo promedio de 36.87% mientras que para el set de datos caracterizados el error fue de 32.54%, con una diferencia de 4% aproximadamente por debajo del set original estandarizado.

Este resultado fué el esperado, ya que el algoritmo analiza el comportamiento de cada uno de los parámetros y de estos se deduce la ecuación característica, por lo tanto, si existe una distorsión o vacíos de información en los parámetros, el algoritmo de aprendizaje no podrá encontrar relaciones adecuadas entre los parámetros de entrada y salida. Al corregir los vacíos de información, modificar los datos y realizar el proceso de caracterización estadística el ajuste de la ecuación es mucho mayor, ya que encuentra el sentido en cada una de las respuestas y analiza con mayor precisión la influencia de todos los parámetros en estudio.

Para analizar la segunda característica establecida para la predicción, la cual corresponde al escalamiento de los datos, se tomaron en cuenta los mismos resultados expresados en la tabla 13, pero, solo tomando en cuenta los sets de datos caracterizados que tienen menor error respecto a la primera característica planteada. En los resultados se observa que el set de datos reales tiene un mejor ajuste con una diferencia de 10% respecto a los datos normalizados y del 12% respecto a los datos estandarizados. Con estos resultados se puede concluir que el mejor set de datos para la predicción con regresión lineal multivariable es el set de datos reales caracterizados.

Con el dataset seleccionado se estudiaron otros parámetros como el error absoluto y el tiempo de cómputo. El error absoluto corresponde al error de la predicción respecto a la cantidad barriles por día de producción, mientras, el tiempo indica la cantidad de segundos que el modelo necesita para encontrar su ecuación característica. Estos resultados se representan en la tabla 14, donde el error absoluto promedio para la arena B2, fue de 111 bbl/día, seguido de la arena D3 con un promedio de error de 93 bpd/día, luego la arena C2 con un promedio de 38 bbl/día y por último la arena E1 con un promedio de 29 bpd/día, para un promedio total de error por macolla de 68 bpd/día.

Al estudiar el tiempo de predicción, se observa que el algoritmo fue bastante rápido en todos los casos, el cual permitió obtener la ecuación característica de un pozo en un tiempo promedio de 0.15 s. Este resultado puede variar de acuerdo con la cantidad de días de producción y la cantidad de parámetros involucrados.

Tabla 14: Errores absolutos y tiempos de predicción por arena en el método de regresión lineal multivariable

Arena	Error absoluto (bpd/d)	t (s)
B2	111	0.16
C2	38	0.12
D3	95	0.19
E1	29	0.14
Promedio	68	0.15

La figura 60 muestra el resultado de la predicción del pozo tipo perteneciente a la arena D3 con el modelo de regresión lineal multivariable, set de datos reales y caracterizados. El color verde oscuro representa la producción real, en color turquesa la predicción del algoritmo en la etapa de entrenamiento y el color naranja representa la predicción en la etapa de prueba. Este pozo en particular presentó un error relativo de 15,52%, y un error absoluto de 25 bpd por día, con un coeficiente de correlación $R^2 = 0,90$, lo cual es un resultado alejado del promedio, por esta razón fue necesario no solo estudiar los resultados por arena, sino por pozo en cada uno de los métodos planteados (Anexo 5: Resultados del modelo de regresión lineal multivariable).

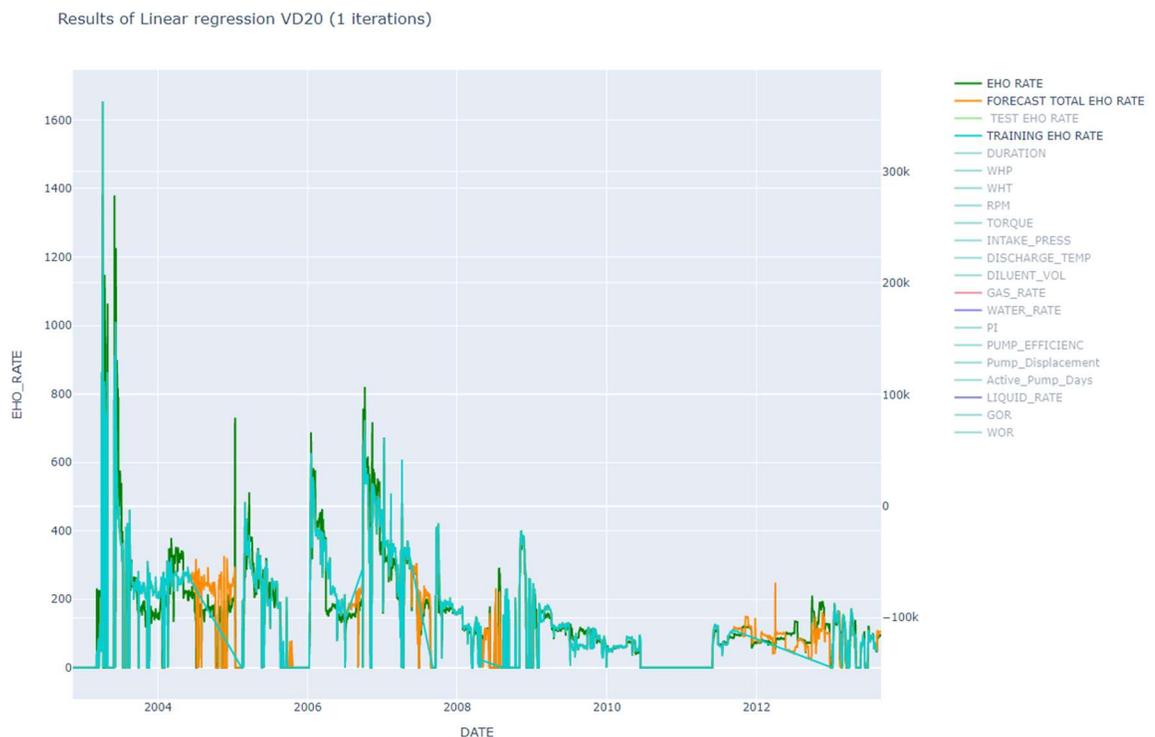


Figura 60: Resultado de la predicción de producción pozo VD20, con el método de regresión lineal multivariable

Este análisis permite concluir que la regresión lineal multivariable es uno de los métodos de *machine learning* que puede ser usado para la predicción de la producción de un pozo, tomando en cuenta todos los parámetros.

4.2.2.- Regresión polinomial

Este modelo es mucho más complejo que el modelo de regresión lineal, ya que el algoritmo devuelve ecuaciones mucho más completas con la capacidad de interrelacionar los parámetros entre sí y generando relaciones polinómicas de acuerdo a cada caso. Para este tipo de modelo al igual que en el anterior se evaluaron los 6 *datasets*, tomando en cuenta ambas características (calidad del datos y escalamiento), sin embargo, se desarrollaron 3 modelos de predicción principales: polinomial de grado 2, polinomial grado 3 y polinomial grado 4. Al inicio se generaron 7 modelos, para analizar el comportamiento de la predicción hasta alcanzar un polinomio de grado 7, sin embargo, no fue posible aplicarlos en los equipos disponibles debido a la capacidad de memoria RAM y procesador instalado.

Todos los modelos antes mencionados se analizaron individualmente para evaluar su respuesta y eficiencia.

4.2.2.1.- Modelo de regresión polinomial de grado 2

El modelo de grado 2 devuelve una ecuación con un promedio de 232 coeficientes, para los cuales el algoritmo procesó aproximadamente 591684 datos por pozo con un coeficiente de correlación promedio de $R^2 = 0.9931$. Este modelo se evaluó en cada uno de los *datasets*, dando como resultado los expresados en la tabla 15.

Tabla 15: Comparación de resultados (error) en los distintos sets de datos para polinomio de grado 2

	Normalizado		Estandarizado		Real	
<i>Arena</i>	<i>Original</i>	<i>Caracterizado</i>	<i>Original</i>	<i>Caracterizado</i>	<i>Original</i>	<i>Caracterizado</i>
B2	19.68%	18.59%	18.21%	19.09%	15.87%	15.71%
C2	30.22%	18.98%	33.26%	14.75%	26.52%	8.79%
D3	30.06%	13.15%	25.25%	12.85%	24.48%	11.62%
E1	9.42%	8.47%	7.82%	6.93%	7.15%	6.47%

Promedio	22.34%	14.80%	21.14%	13.41%	18.51%	10.65%
-----------------	---------------	---------------	---------------	---------------	---------------	---------------

Al analizar este modelo, se observó que los datos originales obtuvieron un error promedio mayor que los datos caracterizados, con una diferencia de un 8% respecto a cada uno de los sets de datos, por lo tanto, se reafirma lo expuesto en el método anterior, en el cual los datos caracterizados obtienen un mejor ajuste en la predicción.

Por otra parte, al comparar los tres sets de datos caracterizados, se observó que los datos normalizados tienen un error relativo promedio de aproximadamente 14.80 %, los estandarizados 13.41 %, mientras que los datos reales presentan un error del 10.65%, con una diferencia de aproximadamente un 3% respecto a los estandarizados y de 4 % respecto a los datos normalizados. Por lo tanto, se observó, que el modelo con menor error para un polinomio de grado 2 se obtiene con el set de datos reales caracterizados.

4.2.2.2.- Modelo de regresión polinomial de grado 3

El modelo de grado 3 devolvió una ecuación con un promedio de 1803 coeficientes, para los cuales el algoritmo procesó aproximadamente 4.431.774 datos por pozo, con un ajuste cuadrático promedio de $R^2 = 0.9988$. Este modelo también se evaluó en cada uno de los *datasets*, dando como resultado los expresados en la tabla 16.

Tabla 16: Comparación de resultados en los distintos sets de datos para polinomio de grado 3

<i>Arena</i>	Normalizado		Estandarizado		Real	
	<i>Original</i>	<i>Caracterizado</i>	<i>Original</i>	<i>Caracterizado</i>	<i>Original</i>	<i>Caracterizado</i>
B2	5449893764.00%	806887621.00%	5371859207.00%	33361213.33%	50.78%	29.36%
C2	1453337042.00%	90820636.37%	48615901.55%	91767872.44%	35.52%	25.18%
D3	1195841431.00%	21565911.68%	21426532.93%	207279311.10%	26.33%	29.16%
E1	8395820.12%	1196851.33%	26919837.06%	32423747.35%	79.02%	37.61%
Promedio	2026867014.28%	230117755.09%	1367205369.64%	91208036.06%	47.91%	30.33%

Al analizar este modelo, se observó que este modelo tanto con el set de datos estandarizados como normalizados no son concluyentes, ya que sus errores relativos superan las 9 cifras, por lo tanto, el único modelo aceptado para este caso es el modelo con el set de datos reales. El error del set de datos originales es de 47.91%, mientras que el caracterizado es de 30.33%, con lo cual se puede concluir que el set de datos ideal para este modelo es el de los datos reales y caracterizados.

4.2.2.3.- Modelo de regresión polinomial de grado 4

El modelo de predicción de grado 4 es el último en este método, y el más complejo, este devolvió una ecuación con un promedio de 11.010 coeficientes, para los cuales el algoritmo procesó aproximadamente 28.053.480 datos por pozo, con un coeficiente de correlación promedio de $R^2 = 0.9995$. Los resultados de este modelo se expresan en la tabla 17.

Tabla 17: Comparación de resultados en los distintos sets de datos para polinomio de grado 4

<i>Arena</i>	Normalizado		Estandarizado		Real	
	<i>Original</i>	<i>Caracterizado</i>	<i>Original</i>	<i>Caracterizado</i>	<i>Original</i>	<i>Caracterizado</i>
B2	16684.29%	2072.83%	3928.67%	975.21%	52.92%	43.88%
C2	195.33%	130.26%	96.32%	169.92%	80.59%	30.02%
D3	2024.33%	370.82%	1102.81%	388.01%	83.59%	45.14%
E1	356.83%	255.35%	102.81%	166.01%	112.37%	41.41%
Promedio	4815.19%	707.32%	1307.65%	424.79%	82.37%	40.11%

Basado en esta tabla, se pudo observar que los valores de error relativo promedio para los sets de datos normalizados y estandarizados son mucho más altos que en el set de datos reales, por lo tanto, estos *datasets* se descartan para este modelo, quedando solo el set de datos reales. El set de datos originales obtuvo en promedio un error del 82.37%, mientras que para el set de datos reales se obtuvo un promedio de

40.11%, por lo tanto, el mejor set de datos para el modelo polinomial de grado 4 es el set de datos reales caracterizados.

4.2.2.4.- Análisis y comparación de los tres modelos polinomiales

Analizando cada uno de los grados se observa que el set de datos óptimo para este modelo de predicción correspondió al set de datos reales caracterizados, ya que con este se obtuvieron los mejores resultados sin importar el grado del polinomio, por lo tanto, se tomaron los resultados de los tres modelos con este set para su análisis y comparación.

Para este análisis se tomaron en cuenta dos parámetros importantes, el coeficiente de correlación R^2 y el error relativo, los cuales se muestran en la tabla 18.

Tabla 18: Comparación de resultados de error relativo y coeficiente de correlación en los métodos de regresión polinomial

Modelo	Error relativo (%)	Coficiente de correlación (r^2)	t (s)
Grado 2	10.65%	0.9931	20.52
Grado 3	30.33%	0.9988	54.24
Grado 4	40.11%	0.9995	217.31

Como se observó en la tabla 18, el modelo con menor error relativo fue el modelo de grado 2, con un error relativo promedio de 10.65%, mientras el de grado 3 tiene un error de 30.33% y el de grado 4 un error de 40.11% con una diferencia respecto al de grado 2 de 20% y 30% respecto al modelo de grado 3. Al estudiar el coeficiente de correlación R^2 , se observó que el modelo con mayor ajuste fue el de grado 4 con un coeficiente de correlación de R^2 de 0.9995, por lo tanto, el de grado 4 debería tener un menor error relativo.

Este fenómeno se estudió con detenimiento y para poder determinar cuál es el modelo óptimo se recurrió al estudio de su error relativo en ambas etapas, tanto en la etapa de entrenamiento como en la etapa de prueba. En la figura 61, se observa un

gráfico comparativo de los errores antes mencionados agrupados según el modelo para la arena B2, la barra de color amarillo representa el error total de la predicción en ambas etapas, la barra de color azul representa el error en la etapa de prueba, mientras el color verde representa el error en la etapa de entrenamiento. El primer paso fue estudiar la etapa de entrenamiento donde el modelo calculó el coeficiente de correlación. En esta etapa el error en el entrenamiento del modelo de grado 4, fue de solo 2.04%, mientras que para los grados inferiores este error va aumentando, siendo el de grado 3 de 3.22% y por último el de grado 2 de 7.91%. Si se observa el error en la etapa de prueba ocurrió todo lo contrario, el menor error se observó en el modelo de grado 2, con un error relativo de 27.11% y a medida que aumenta el grado del polinomio, este error aumentó, siendo para el modelo de grado 3 un error de 90.33% y para el modelo de grado 4 fue de 141.32 %.

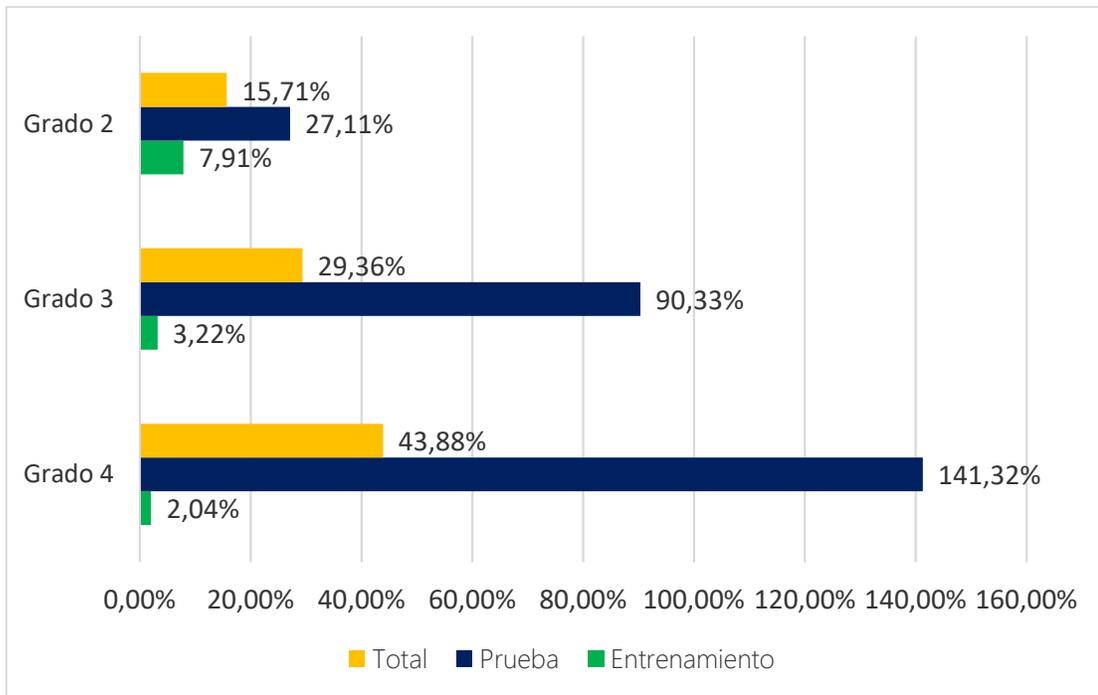


Figura 61: Estudio del modelo de regresión polinomial en la arena B2

Basados en este análisis se observó que entre mayor es el grado del polinomio, el modelo tiende a ajustarse más a los datos de entrenamiento, pero al momento de enfrentarse a un set de datos nuevos y tratar de predecir su comportamiento, el

algoritmo se comporta de forma ineficiente, este tipo de comportamiento se conoce como sobreajuste.

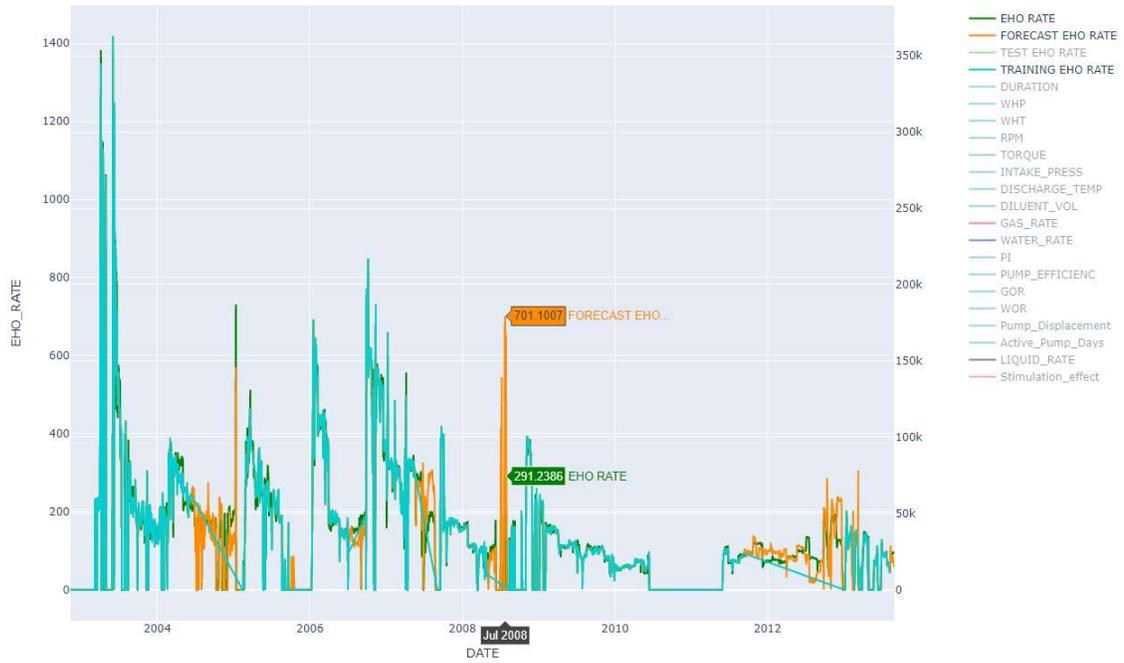
Por esta razón, a pesar de que el modelo de grado 4 tenía un mayor coeficiente de correlación, el modelo de grado 2 se ajustó mejor a los objetivos deseados.

Otro factor importante a estudiar es el tiempo de ejecución. En la tabla 27 se muestra que a medida que aumenta el grado del polinomio, el tiempo aumenta, sin embargo, el tiempo de predicción para los 3 métodos es bajo: siendo el mayor de 217.31 segundos, lo cual es aproximadamente 4 minutos para una predicción compleja, a este le sigue el de grado 3, con un tiempo de ejecución de 54.24 segundos y por último el de grado 2 con un tiempo aproximado de 20.52 segundos para la predicción de un pozo.

En la figura 62 se muestran los resultados del pozo VD20 en los tres modelos de regresión polinomial con el set de datos reales caracterizados, el color verde oscuro representa la producción real, en color turquesa la predicción del algoritmo en la etapa de entrenamiento y el color naranja representa la predicción en la etapa de prueba. Como se observa en la figura 62a, la predicción del algoritmo de grado 2 es muy similar a los datos reales, sin embargo, existen anomalías como la que se presenta en el mes de julio de 2008, donde la diferencia de la producción entre la predicción y el real es de aproximadamente 400 bpd. En la figura 62b se observa el modelo de grado 3, con grandes picos y diferencias entre los datos reales y la predicción, sin embargo, para el mismo mes de julio de 2008, la diferencia entre ambos datos es de hasta 7.600 bpd por día, patrón que se repite en la predicción de grado 4, como se observa en la figura 63, donde la diferencia entre la predicción y la tasa real es significativa, e incluso para el mismo mes antes expuesto, la diferencia entre la predicción y el valor real supera los 14.000 bpd / día. Este patrón observado en los tres modelos es de gran ayuda al momento de estudiar las anomalías en los datos y posibles problemas en el sistema pozo yacimiento.

Cada uno de los pozos presenta resultados diferentes para cada uno de los grados, por lo tanto, es necesario estudiar cada uno con detalle para definir el mejor modelo de este grupo (Anexo 6: Resultados del modelo de regresión polinomial).

Results of Polinomial regression VD20 (Degree 2)



Results of Polinomial regression VD20 (Degree 3)



Figura 62: Resulta del modelo de regresión polinomial del pozo VD20. (a) Regresión polinomial grado 2 (b) regresión polinomial grado grado?

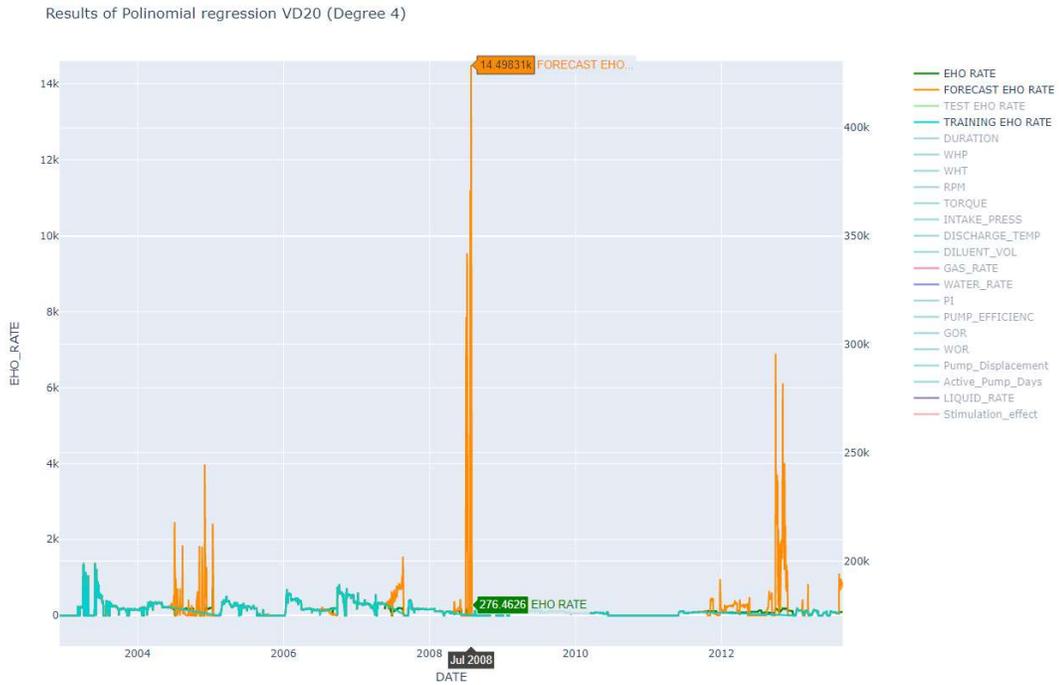


Figura 63: Resultado del modelo de regresión polinomial grado 4pozo VD20

4.2.3.- Redes neuronales

El modelo de predicción de redes neuronales es el más complejo de los tres modelos desarrollados en este TEG. Este tipo de modelos no devuelve una ecuación característica, sino que desarrolla una estructura de cálculo compleja con funciones y capas predeterminadas por el usuario. Este proceso se evaluó en dos etapas. En la primera etapa se evaluaron 10 pozos (5 pozos pertenecientes al ambiente sedimentario fluvial y 5 pozos pertenecientes al ambiente sedimentario deltaico) tomando en cuenta dos características: la calidad del dato (caracterizados y sin caracterizar) y la función de pérdida (MAE y MSE). Los resultados obtenidos para los pozos del deltaico son los expuestos en la tabla 19, donde se observan los mejores casos por pozo resaltados en color verde.

Tabla 19: Resultados de la etapa 1 en el ambiente sedimentario deltaico

Deltaico				
Nombre del pozo	Set	Funcion de pérdida	Error absoluto Promedio (bpd)	Error relativo promedio (%)
VD03	Caracterizado	MAE	100.47	18.83
VD03	Original	MAE	117.18	22.12
VD03	Caracterizado	MSE	128.92	22.47
VD03	Original	MSE	151.01	25.65
VD05	Caracterizado	MSE	122.46	27.38
VD05	Caracterizado	MAE	115.35	28.39
VD05	Original	MSE	122.01	29.47
VD05	Original	MAE	125.86	30.46
VD14	Caracterizado	MAE	85.27	18.96
VD14	Original	MAE	81.68	19.1
VD14	Caracterizado	MSE	93.53	20.17
VD14	Original	MSE	91.28	22.54
VD21	Original	MSE	153.04	23.24
VD21	Caracterizado	MAE	142.23	23.3
VD21	Original	MAE	152.97	25.15
VD21	Caracterizado	MSE	154.59	25.67
VD40	Caracterizado	MAE	71.23	24.74
VD40	Original	MAE	75.99	28.47
VD40	Original	MSE	63.75	29.09
VD40	Caracterizado	MSE	73.48	34.91

El set de datos caracterizados tiene un mejor ajuste y menor error relativo en cuatro de los cinco pozos, lo que correspondió a un 80% de la muestra seleccionada, mientras que en un pozo el mejor set de datos es el set no caracterizado con un 20%. Respecto a la función de pérdida se tiene que el menor error relativo promedio se obtiene con la función de error absoluto promedio (MAE, por sus siglas en ingles), con un total de tres pozos respecto a los cinco en estudio, lo que corresponde a un 60% de la muestra, mientras la función de pérdida de error cuadrático promedio (MSE, por sus siglas en inglés) resulta ser la mejor para dos pozos, con un porcentaje del 40% para el total de la muestra.

Al analizar los errores en el yacimiento fluvial en la tabla 20, se observó que el comportamiento de los resultados es similar al antes expuesto, en donde el set de datos caracterizados, tienen un mejor ajuste en el 80% de los pozos de la muestra, mientras el original solo en el 20%. De igual forma, se estudió la función de pérdida donde se obtuvo que la función MAE tuvo un menor error relativo promedio en cuatro de los cinco pozos, lo que corresponde al 80% de los pozos, mientras la función MSE obtuvo un menor error relativo solo en un pozo, correspondiente a un 20%.

Tabla 20: Resultados de la etapa 1 en el ambiente sedimentario fluvial

Fluvial				
Nombre del pozo	Set	Funcion de pérdida	Error absoluto Promedio (bpd)	Error relativo promedio (%)
VD02	Original	MAE	165.65	54.33
VD02	Caracterizado	MAE	187.33	67.41
VD02	Original	MSE	166.68	68.03
VD02	Caracterizado	MSE	179.09	71.4
VD07	Caracterizado	MAE	149.79	22.18
VD07	Caracterizado	MSE	152.33	24.29
VD07	Original	MAE	154.33	25.06
VD07	Original	MSE	163.08	28.16
VD10	Caracterizado	MAE	41.04	13.23
VD10	Original	MSE	47.97	14.57
VD10	Original	MAE	50.07	16.14
VD10	Caracterizado	MSE	47.42	16.72
VD18	Caracterizado	MSE	1243.76	225.17
VD18	Caracterizado	MAE	1239.76	228.4
VD18	Original	MSE	1254.48	234.38
VD18	Original	MAE	1348.67	249.87
VD24	Caracterizado	MAE	123.32	25.52
VD24	Original	MSE	132.11	29.61
VD24	Caracterizado	MSE	133.62	31.83
VD24	Original	MAE	138.88	34.28

En la figura 64 se muestran los resultados de este modelo agrupados respecto a la función de pérdida y el tipo de *dataset*, donde se demuestra para esta etapa que el *dataset* con mejores resultados, fue el set con datos caracterizados con un 80%, mientras respecto a la función de pérdida el resultado con mejor ajuste se obtuvo con

la función MAE con un porcentaje del 70%. Sin embargo, adicional a estos porcentajes obtenidos, se estudiaron ambas tablas con detalle (tablas 19 y 20), donde se observaron que en aquellos pozos en los cuales el set de características seleccionadas (caracterizado / MAE) no dio los mejores resultados; este modelo es el segundo más cerca, con diferencias pequeñas en el error relativo.

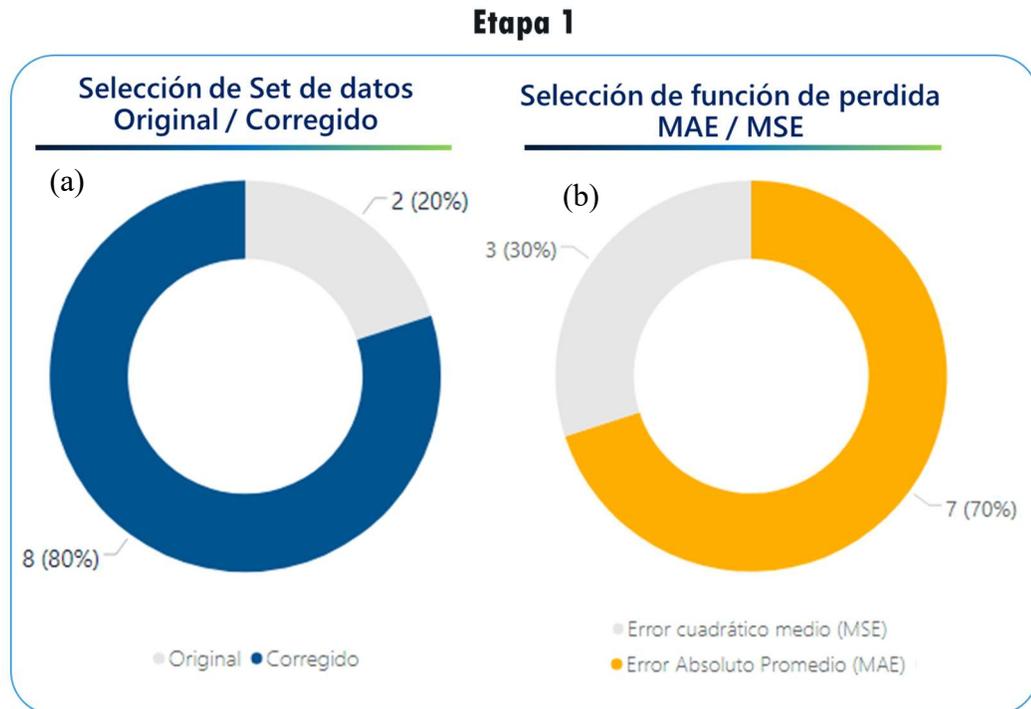


Figura 64: Resultados de la etapa 1 (Metodología redes neuronales) (a) Selección del set datos optimo (b) Selección de función de pérdida

Para la etapa 2 del modelo se evaluó el escalamiento tomando en cuenta los resultados de la etapa 1, donde la función de pérdida ideal para el modelo fue MAE y el set de datos caracterizados. Para esta etapa se evaluaron los mismos 10 pozos, solo que, comparando el resultado de los datos reales caracterizados, con los normalizados y estandarizados caracterizados. En la figura 65, se observan los resultados finales del modelo, donde seis (6) de los 10 pozos en estudio, obtuvieron mejores resultados con el set de datos normalizados, lo que equivale a un 60% de la muestra, mientras que para los otros dos sets de datos existió una igualdad porcentual del 20%. Por lo tanto, las

características adecuadas para este modelo resultaron ser: set de datos normalizados caracterizados con función de pérdida MAE.

Etapa 2

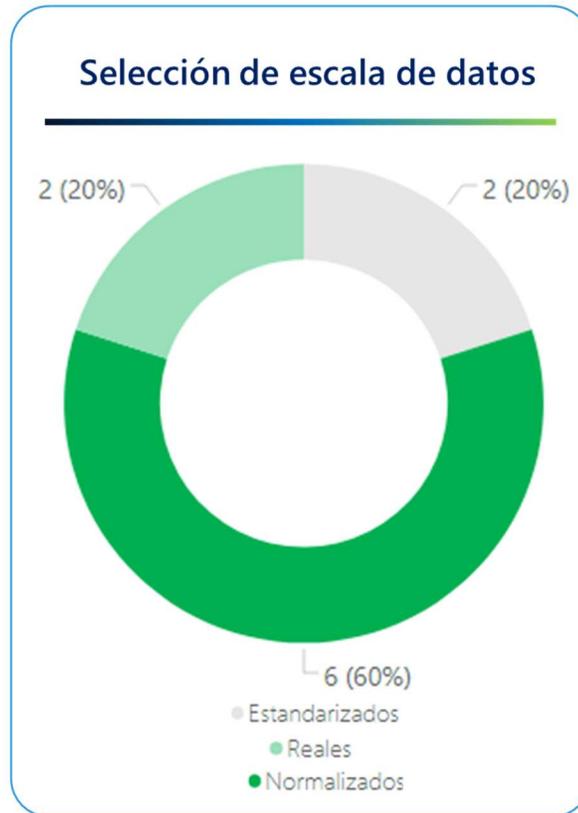


Figura 65: Resultados de la etapa 2 (Metodología redes neuronales) (a) Selección del escalamiento

A diferencia de los modelos anteriores, el mejor ajuste con este modelo se encuentra con los datos normalizados, este resultado es el esperado debido a la complejidad del modelo y a la cantidad de cálculos internos que realiza el algoritmo, lo que permite mayor capacidad de cómputo y menor tiempo al tener menor distorsión en la escala de datos.

Al definir las mejores características para el modelo de redes neuronales, el algoritmo se aplicó a los 40 pozos y éstos se filtraron por arena, con la finalidad de estudiar el comportamiento del modelo. El código con el algoritmo se ejecutó de manera remota, usando los servidores de Google, mediante la plataforma Google

Colab, razón por la cual se desconoce la información técnica de la computadora que ejecutó dicho algoritmo. Una vez agrupados los pozos por arena se determinó el error relativo promedio, error absoluto y el tiempo de predicción, obteniendo los siguientes resultados mostrados en la tabla 21, donde la arena D3 tuvo el mayor error relativo con 34.07%, un error absoluto de 129 bpd y el mayor tiempo de predicción de aproximadamente 32 minutos (describir al principio el computador, procesadores, etc., le siguió la arena B2 con 26.23 % de error relativo, un error absoluto de 108 bpd y con un tiempo de predicción de 38 minutos para el entrenamiento del modelo, por último, las arenas C2 y E1, con un error relativo de 16.18 % y 14.16% respectivamente, un error absoluto menor a 50 bpd, y con un tiempo de predicción aproximadamente de 20 minutos. Las arenas con menor cantidad de datos por pozo obtuvieron resultados en menor tiempo.

Tabla 21: Resultados del modelo de redes neuronales por arena

Arena	Error relativo (%)	Error absoluto (bpd)	Tiempo (s)
B2	26.23%	108	2321.48
C2	16.18%	46	1181.69
D3	34.07%	129	1925.67
E1	14.16%	35	1200.27
Promedio	22.66%	79.457175	1657.27833

En la figura 66 se muestran los resultados del modelo de redes neuronales en el pozo VD20 perteneciente a la arena D3, donde el color verde representa la producción real de petróleo, mientras el color rojo representa la predicción del modelo. La arquitectura de esta red presentó tres capas ocultas compuestas cada una por 50 neuronas, donde su tiempo de entrenamiento fue de aproximadamente 40 minutos., este modelo en particular tuvo un error relativo de 20.48% y un error absoluto de 28 bpd, como se observó estos errores son mucho menores que los que tiene la arena D3, por esta razón fue necesario comparar los resultados en cada uno de los modelos por pozo

para determinar cuál de estos describe mejor su comportamiento (Anexo 7: Resultados del modelo de redes neuronales).

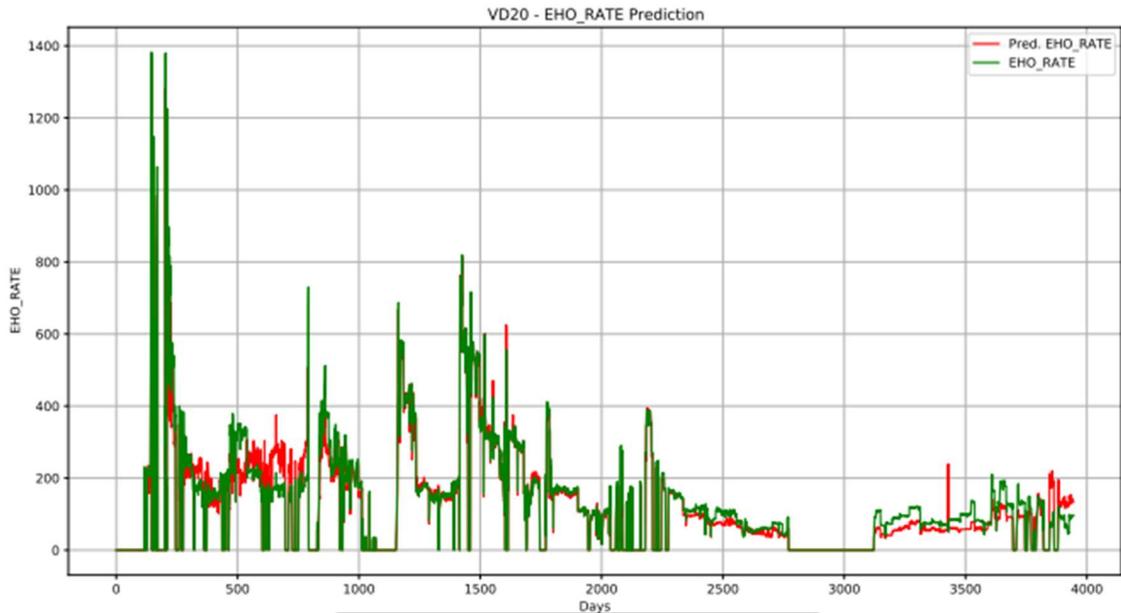


Figura 66: Resultado del modelo de redes neuronales en el pozo VD20

4.2.4.- Importancia de la caracterización estadística

Como respuesta del análisis realizado en todos los modelos se observó que la calidad del dato juega un papel fundamental en la aplicación y factibilidad de este tipo de metodologías.

La figura 67 muestra los resultados promedios de cada modelo respecto a la primera característica en estudio (calidad del dato). Las barras en color azul representan los sets de datos reales, mientras las barras de color verde representan los sets de datos caracterizados. En el caso de las redes neuronales, el error relativo para los datos crudos sin caracterizar superó por casi un 20% al resultado obtenido con el set de datos caracterizados. Este patrón se repite en todos los modelos, sin embargo, la diferencia de error relativo entre ambos dataset osciló entre 8 y 20% de acuerdo con cada modelo.

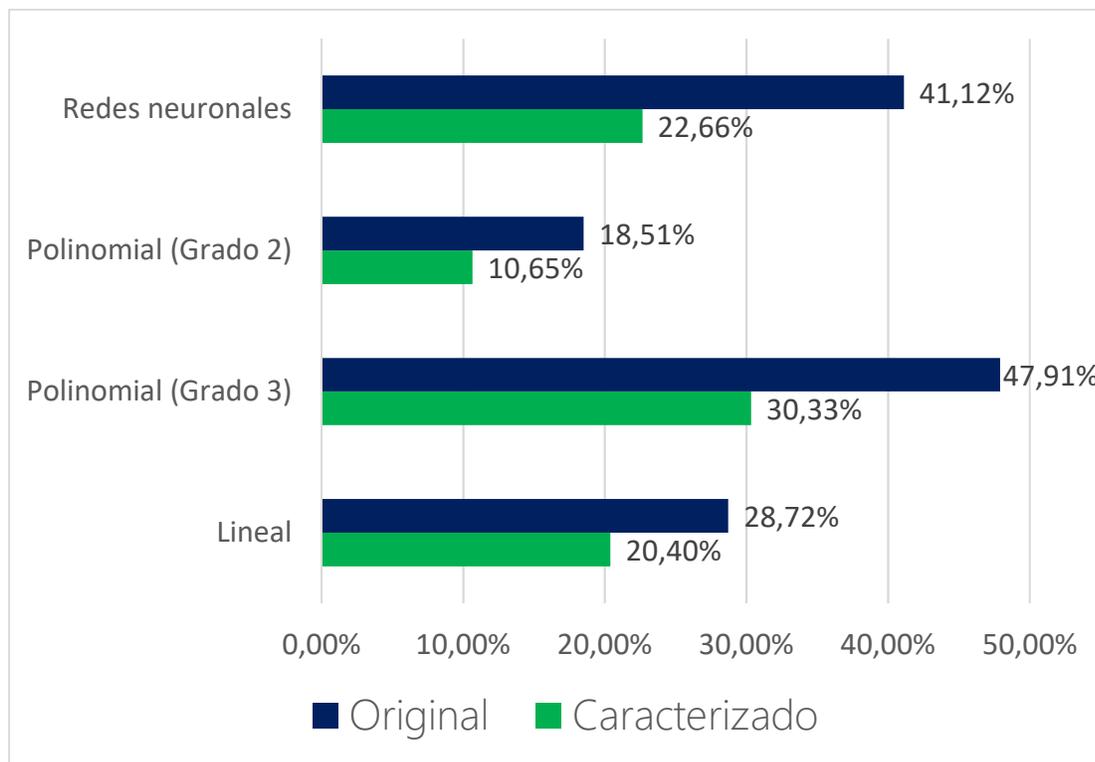


Figura 67: Análisis de resultados de los tres modelos de acuerdo con la calidad del dato

El estudio demostró que los datos fuera de rango y datos nulos aumentan la incertidumbre en las predicciones de los modelos. En la evaluación de la base de datos del historial de producción se encontraron gran número de valores erróneos como: datos de presión de fondo o temperatura por debajo de la atmosférica o presiones y temperaturas negativas, datos nulos o faltantes, eficiencias de la bomba mayores a uno, entre otros. Con base en la teoría explicada acerca de las metodologías de *data science* se sabe que mientras existan datos erróneos, faltantes, con mucha dispersión o poco sentido no se obtendrá un aprendizaje exitoso, ya que el algoritmo procesa estos datos, aprende y obtiene resultados. Si entrenamos a nuestro algoritmo con una base de datos con muchos errores, este no encontrará la ecuación representativa del fenómeno, por lo tanto, al cambiar a los datos reales o mejorar la calidad del dato, el algoritmo no será capaz de predecir, ya que recibió un mal entrenamiento. Por lo tanto, se determinó que el set de datos idóneo para esta investigación fue el set de datos caracterizados.

4.2.5.- Importancia de la normalización

La normalización fue un proceso que permitió optimizar el tiempo de cómputo, sin embargo, los resultados obtenidos al aplicar este tipo de metodologías generaron mayores errores relativos en todos los modelos.

Para este análisis se tomó en cuenta solamente el set de datos caracterizados que fue el de menor error respecto a la calidad del dato. En la figura 68 se muestran los resultados finales de cada modelo, en color amarillo se representa el error relativo al estandarizar los datos, en color azul los resultados al normalizar los datos y, por último, en color verde los resultados con el set de datos reales. Siendo el set de datos reales el de menor error y con tiempos de respuesta bastante rápidos.

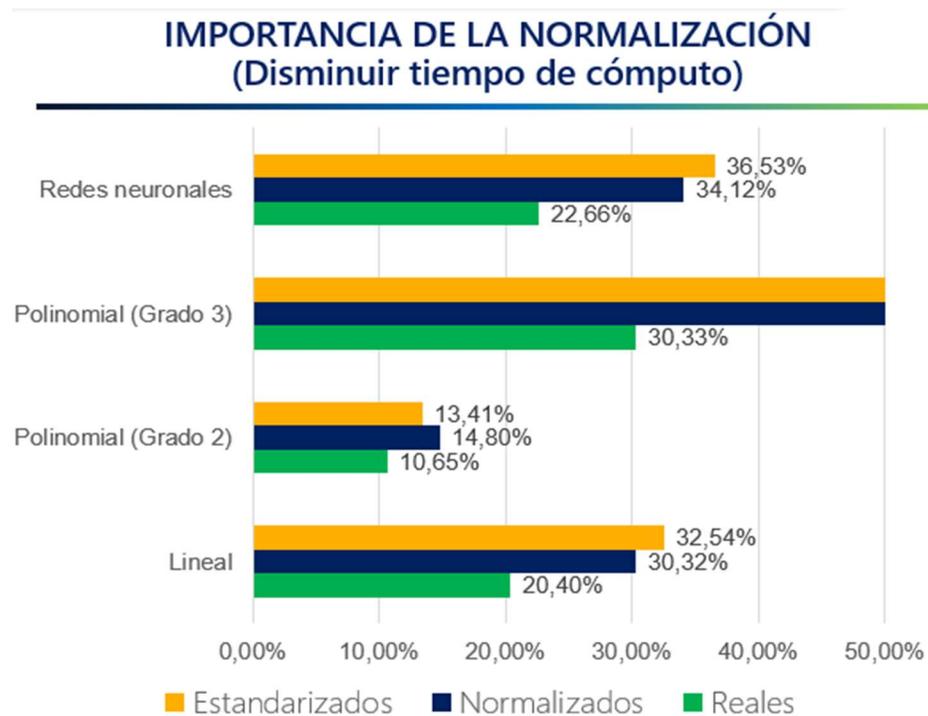


Figura 68: Análisis de resultados de los tres modelos respecto a la normalización

4.2.6.- Selección del modelo final de predicción

Una vez desarrollados y analizados los resultados en los tres modelos, los mismos fueron comparados con la finalidad de definir el modelo óptimo para cada pozo. En la tabla 22, se muestran los resultados agrupados por cada modelo, donde se observó que 27 de los 40 pozos en estudio tienen un mejor ajuste con el modelo de regresión polinomial de grado 2, lo que corresponde a un 67.50% de la muestra, 8 pozos presentaron un mejor ajuste con el modelo de redes neuronales lo que correspondió a un 20%, mientras que 4 pozos presentaron un mejor ajuste con el modelo de regresión lineal lo que representó solo un 10% de los casos y por último, 1 pozo presentó mejor ajuste por regresión polinomial de grado 3 (Anexo 8: Comparación numérica de resultados entre modelos).

Tabla 22: Selección del mejor modelo por pozo

Modelo	Cant. Pozos	%
Lineal	4	10.00%
P. Grado 2	27	67.50%
P. Grado 3	1	2.50%
Red Neuronal	8	20.00%
Total	40	100.00%

Las tablas 23 y 24 muestran una comparación del error relativo y el error absoluto de los tres modelos con mayor número de casos, donde el menor error relativo y absoluto en todas las arenas y respecto a la macolla se consiguió con el modelo de regresión polinomial de grado 2, obteniendo un error relativo de aproximadamente 10.65 %, con una diferencia de 10 a 12% respecto a los otros dos modelos, mientras el error absoluto es de aproximadamente 40 bpd/día con una diferencia de 28 bpd/día respecto al modelo lineal y de 39 bpd/día respecto al modelo de redes neuronales.

Tabla 23: Comparación del error relativo en los tres modelos principales

Arena	ANN	Lineal	Polinomial Grado 2
B2	26.23%	23.62%	15.71%
C2	16.18%	15.29%	8.79%
D3	34.07%	27.28%	11.62%
E1	14.16%	15.41%	6.47%
Macolla	22.66%	20.40%	10.65%

Tabla 24: Comparación del error absoluto (bbls/día) en los tres modelos principales

Arena	ANN	Lineal	Polinomial Grado 2
B2	108	111	70
C2	46	38	21
D3	129	95	49
E1	35	29	21
Macolla	79	68	40

En la tabla 25, se muestra la comparación de los tres modelos respecto al tiempo de entrenamiento, donde el modelo de regresión lineal es el modelo más rápido con un tiempo de 0.15 segundos, sin embargo, el modelo polinomial de grado 2 que fue el modelo con menor error, se consideró un modelo que obtiene una rápida respuesta, con

un tiempo de entrenamiento aproximado de 19.36 s, y por último el modelo de redes neuronales con un tiempo 1657.28 segundos en promedio.

Tabla 25: Comparación del tiempo de entrenamiento (segundos) en los tres modelos principales

Arena	ANN	Lineal	Polinomial Grado 2
B2	2321.48	0.16	25.40
C2	1181.69	0.12	15.66
D3	1925.67	0.19	21.89
E1	1200.27	0.14	14.49
Macolla	1657.28	0.15	19.36

Estas comparaciones permitieron determinar que el mejor modelo de predicción de manera general fue el modelo de regresión polinomial, ya que tuvo menor error relativo, menor error absoluto y fue el modelo con mejor ajuste en la mayoría de los pozos. Por otra parte, a pesar de no ser el modelo con menor tiempo de entrenamiento, se consideró que consigue resultados mucho más precisos en poco tiempo (19 segundos), bastante rápido en comparación con las metodologías actuales de análisis de datos en el campo.

4.3.- Significancia estadística

La significancia estadística es uno de los parámetros principales para determinar la importancia de un atributo en la predicción, entre menor sea el valor de p, indica que el parámetro tiene una gran influencia en la ecuación característica y por lo tanto en el parámetro de salida y en los resultados obtenidos.

El análisis de la significancia estadística fue fundamental para el posterior análisis de los pozos, ya que, este parámetro permitió definir la influencia de los mecanismos de empuje y operatividad en cada uno.

En la figura 69 se muestra la significancia estadística (valor p) de cada uno de los atributos para la arena B2, incluyendo las bombas, donde se observó que los atributos con una fuerte evidencia de correlación ($p < 0.0001$) correspondieron a: la tasa de gas, el índice de productividad, la presión de entrada de la bomba y las RPM de la bomba. Adicionalmente existen tres modelos de bombas con fuerte evidencia de correlación, como lo son los modelos 200TP1800, el 165-1350 y el 300TP1800. Por otro lado, existen otros atributos con una evidencia moderada de correlación ($p < 0.05$), como: la temperatura de descarga, la temperatura de cabezal, los efectos de la estimulación, la eficiencia de la bomba, la presión de cabezal y el torque. Los otros parámetros en esta arena no mostraron tener una significancia estadística tan importante, sin embargo, estos resultados pueden variar respecto a cada uno de los pozos en la misma arena (Anexo 8.6. Resultados de significancia estadística de los pozos en la arena B2). El análisis de la significancia estadística de la arena determinó que uno de los mecanismos de empuje presentes en el yacimiento y con mayor relevancia es el gas. Este resultado fue de esperarse por la distribución de los yacimientos dentro del campo y el ambiente sedimentario al cual pertenece.

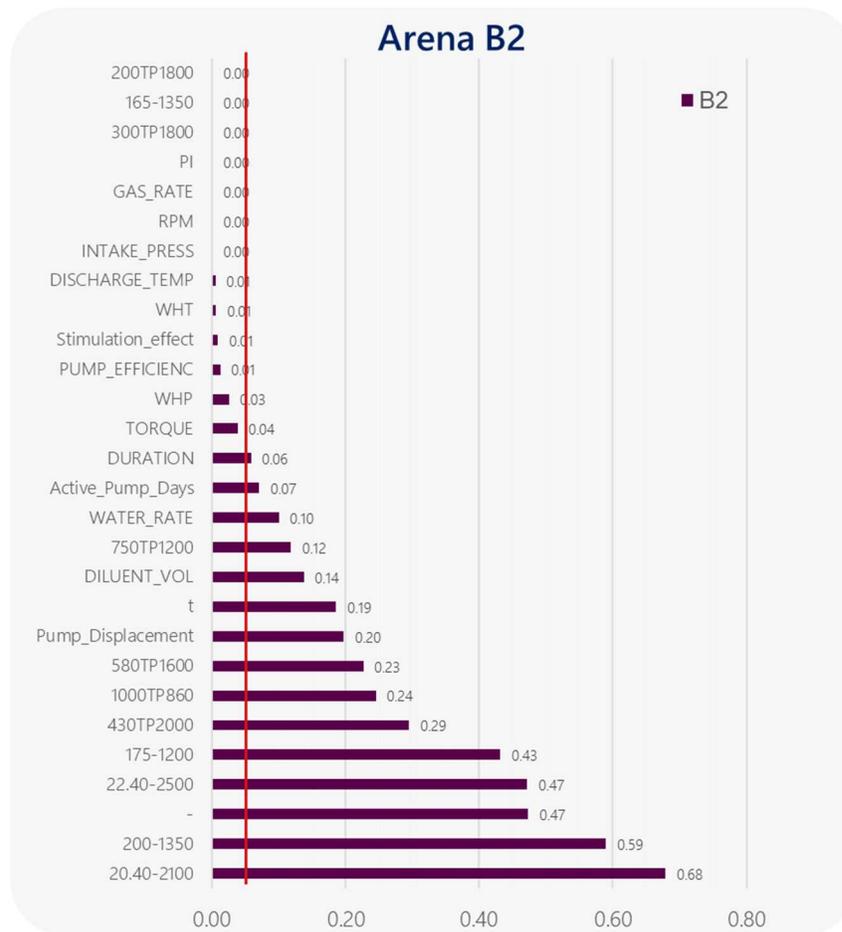


Figura 69: Significancia estadística de los atributos en la arena B2

Por otra parte, la significancia estadística de cada una de las bombas es un factor determinante para el estudio de su eficiencia el cual se analiza en la siguiente sección.

En la figura 70, se muestra la significancia estadística de los atributos en la arena C2, donde los atributos con una fuerte evidencia de correlación correspondieron a: la tasa de gas, las RPM, y los modelos de bombas, 60TP1300, 98-1600 y 20-1350. Hay otros parámetros que mostraron evidencia de correlación moderada como: índice de productividad, la presión de cabezal, la eficiencia de la bomba, la presión de entrada de la bomba, la temperatura de cabezal, la duración, la tasa de agua, la cantidad de días activos de la bomba, y el tiempo, sin embargo, estos resultados varían respecto a cada pozo (Anexo 8.7. Resultados de significancia estadística de los pozos en la arena C2).

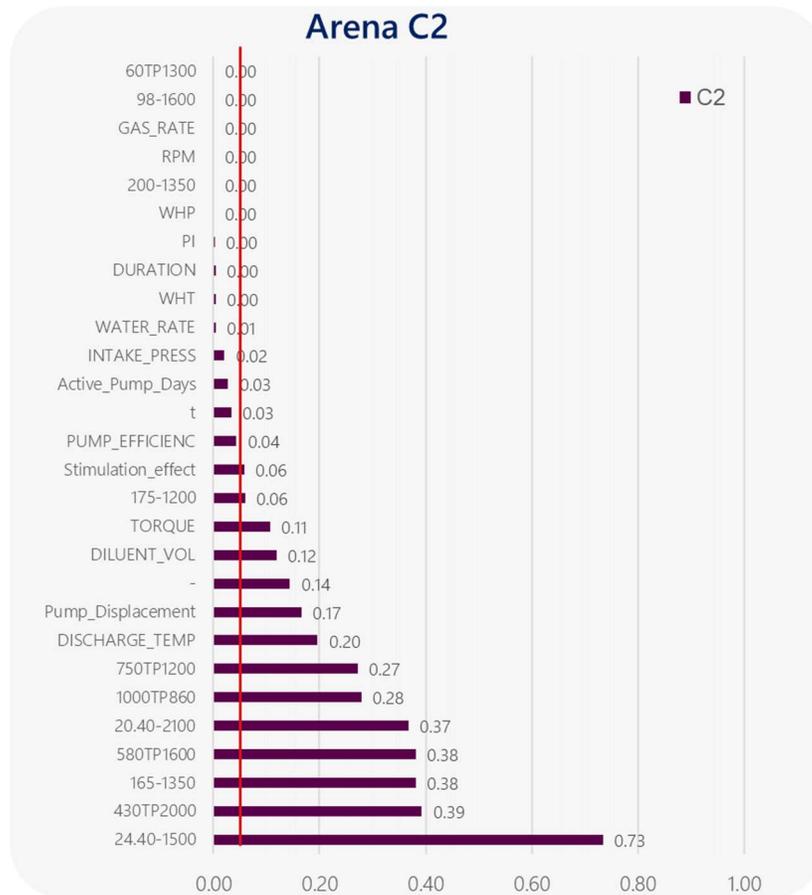


Figura 70: Significancia estadística de los atributos en la arena C2

Estos resultados permitieron definir nuevamente el mecanismo de empuje presente en este yacimiento (empuje por gas), sin embargo, se observó que para algunos pozos del yacimiento C2, la tasa de agua tiene una gran importancia en la producción.

En la figura 71, se muestra el resultado de la significancia estadística de la arena D3, donde los atributos con fuerte evidencia de correlación correspondieron a: la tasa de gas, la eficiencia de la bomba, las RPM, y los modelos de las bombas, 130-1200, 22-40-2500, 24-40-1500, 750TP1200, 28-40-500, mientras los atributos con moderada evidencia de correlación fueron: la presión de cabezal, el tiempo, la temperatura de descarga y el modelo de bombas 1000TP860. De igual forma como en los casos anteriores, la importancia de estos atributos puede variar de acuerdo con el pozo en estudio.

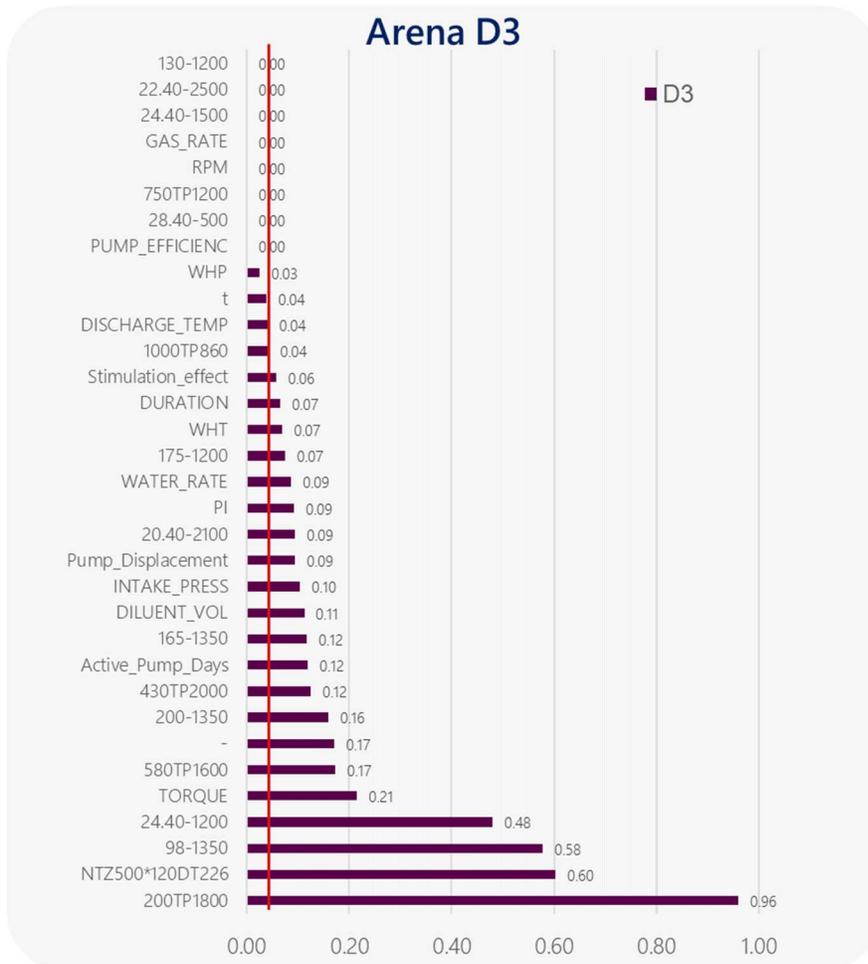


Figura 71: Significancia estadística de los atributos en la arena D3

Por último, se estudió la significancia estadística de los atributos en la arena E1, donde se observó que la mayor cantidad de atributos con evidencia de correlación, fueron: la tasa de gas, la presión de entrada de la bomba, el índice de productividad, la eficiencia de la bomba, y los modelos de bomba 400TP1350, 20.40-2100, 165-1350, 1000TP860, como se muestra en la figura 72. Por otro lado, se tiene que los atributos con moderada evidencia de correlación resultaron ser la presión de cabezal, el tiempo, el volumen de diluyente, el torque, y los modelos de bomba, 200-1350, 430TP2000, 175-1200 y el modelo 750TP1200.

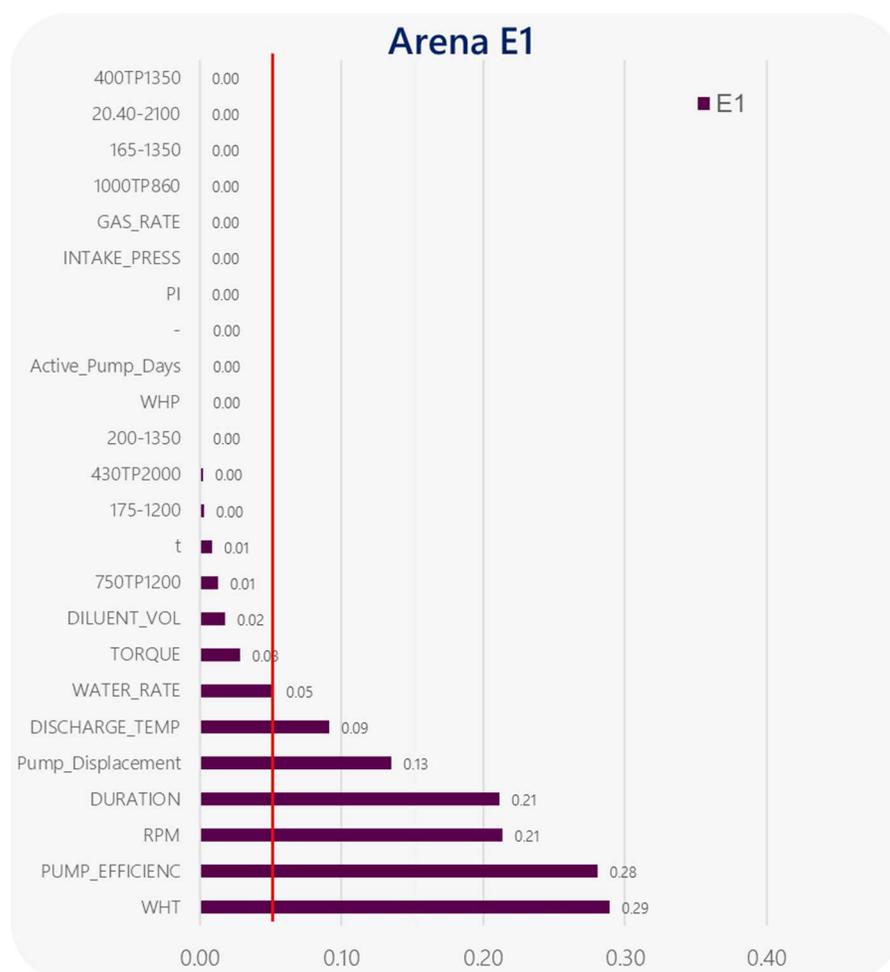


Figura 72: Significancia estadística de los atributos en la arena E1

Ahora, si se comparan las arenas de acuerdo con el ambiente sedimentario al que pertenecen, se pudo observar que existen atributos en ambas arenas con fuerte evidencia de correlación. En la tabla 26, se encuentra la comparación de los atributos para los yacimientos del deltaico (B2 y C2), sin incluir los modelos de bomba. En estos resultados se observó que en ambas arenas, los parámetros de tasa de gas, presión de cabezal, eficiencia de la bomba, índice de productividad, RPM, presión de entrada de la bomba y temperatura de cabezal, son importantes y tienen una alta correlación con el atributo de salida (tasa de petróleo), sin embargo, hay otros parámetros como la producción de agua, la cual tuvo mayor impacto en la arena C2, un caso similar ocurre

con los efectos de la estimulación, donde se observó que en la arena B2, estos generan un mayor impacto en la producción. Otros atributos como la duración y el tiempo son parámetros que tienen una mayor correlación en la arena C2, sin embargo, la temperatura de descarga tiene un mayor efecto en la arena B2.

Tabla 26: Comparación de significancia estadística para las arenas del ambiente deltaico

	B2	C2
GAS_RATE	0.0000	0.0000
WHP	0.0253	0.0005
PUMP_EFFICIENC	0.0121	0.0427
PI	0.0000	0.0029
RPM	0.0000	0.0000
INTAKE_PRESS	0.0001	0.0210
WHT	0.0062	0.0039
DURATION	0.0581	0.0038
WATER_RATE	0.1004	0.0053
Active_Pump_Days	0.0704	0.0280
t	0.1853	0.0340
Stimulation_effect	0.0091	0.0586
TORQUE	0.0381	0.1065
DILUENT_VOL	0.1373	0.1193
Pump_Displacement	0.1972	0.1669
DISCHARGE_TEMP	0.0053	0.1962

En la tabla 27, se observa la comparación de los atributos para el yacimiento fluvial, en el que se encuentran las arenas D3 y E1: En este caso observaron solo tres atributos con importante evidencia de correlación, los cuales correspondieron a la tasa de gas, la presión de cabezal y la eficiencia de la bomba. La arena D3, es una de las arenas con menor evidencia de correlación en sus atributos a diferencia de la arena E1, sin embargo, se observa que la temperatura de descarga tiene una moderada evidencia de correlación. Otro parámetro importante en esta arena es el efecto de la estimulación, que a pesar de no estar señalado en color verde en la tabla 27, su valor está muy cercano a 0.0500, lo que indica que este atributo es importante.

Para la arena E1, existe una fuerte evidencia que el agua, la presión de entrada de la bomba y el índice de productividad tienen una gran influencia en la tasa de petróleo. Adicionalmente se observó que en esta arena no se ha realizado ningún trabajo de estimulación, por lo tanto, el algoritmo no puede determinar su significancia estadística (la importancia de estos atributos puede variar entre los pozos).

Tabla 27: Comparación de significancia estadística para las arenas del ambiente fluvial

	D3	E1
GAS_RATE	0.0000	0.0000
WHP	0.0250	0.0059
PUMP_EFFICIENC	0.0001	0.0001
PI	0.0926	0.0000
RPM	0.0000	0.0868
INTAKE_PRESS	0.1033	0.0000
WHT	0.0696	0.1271
DURATION	0.0650	0.0725
WATER_RATE	0.0857	0.0258
Active_Pump_Days	0.1192	0.1626
t	0.0379	0.0596
Stimulation_effect	0.0570	
TORQUE	0.2145	0.0234
DILUENT_VOL	0.1122	0.0628
Pump_Displacement	0.0942	0.0454
DISCHARGE_TEMP	0.0402	0.2566

En la siguiente sección se analizarán todos los parámetros relacionados con el funcionamiento de las BCP y los resultados obtenidos por los algoritmos de clasificación desarrollados en la metodología.

4.4.- Análisis del modelado del comportamiento de las BCP

El desarrollo de los algoritmos para el modelo de análisis de las BCP, se generó en base a tres herramientas de clasificación: KNN, *support vector machine* y árbol de decisiones. Las herramientas de KNN y *support vector machine* no arrojaron ningún

resultado concluyente para nuestra investigación debido a la complejidad de los datos. Este tipo de algoritmos de clasificación puede usarse para clasificar grupos de datos definidos en base a algunas características, sin embargo, ninguno de los dos algoritmos fue capaz de detectar anomalías con las variables que se presentaron.

El modelo que mejor se adaptó a los requerimientos fue el modelo de árbol de decisiones, el cual generó una jerarquía interconectada entre los parámetros de estudio y se obtuvo como resultado 24 casos principales, los cuales, permiten determinar el estado de la bomba en un periodo de producción determinado.

Todos los casos son representados a través de una gráfica interactiva como se muestra en la figura 73. La figura 73a muestra la gráfica resultante de la bomba 1 del pozo tipo, en la esquina superior derecha se encuentran los resultados hallados en la primera fase de predicción (*Prediction e Ideal rate real*) los cuales pueden habilitarse o inhabilitarse según los requerimientos del usuario para un correcto análisis de datos. La línea de color azul representa el resultado del modelado de predicción, mientras la línea naranja representa la tasa ideal de la bomba a condiciones de subsuelo.

Por otro lado, en la figura 73b, muestra el resultado del análisis a través de la metodología de árbol de decisiones, el lado derecho de la gráfica muestra una escala de color que varía desde el color verde para indicar un buen funcionamiento de la bomba, color amarillo y naranja para mostrar un comportamiento regular y las tonalidades de color rojo para mostrar un mal funcionamiento de la bomba o alguna anomalía en el sistema.

VD20: Pump analysis 430TP2000



VD20: Pump analysis 430TP2000

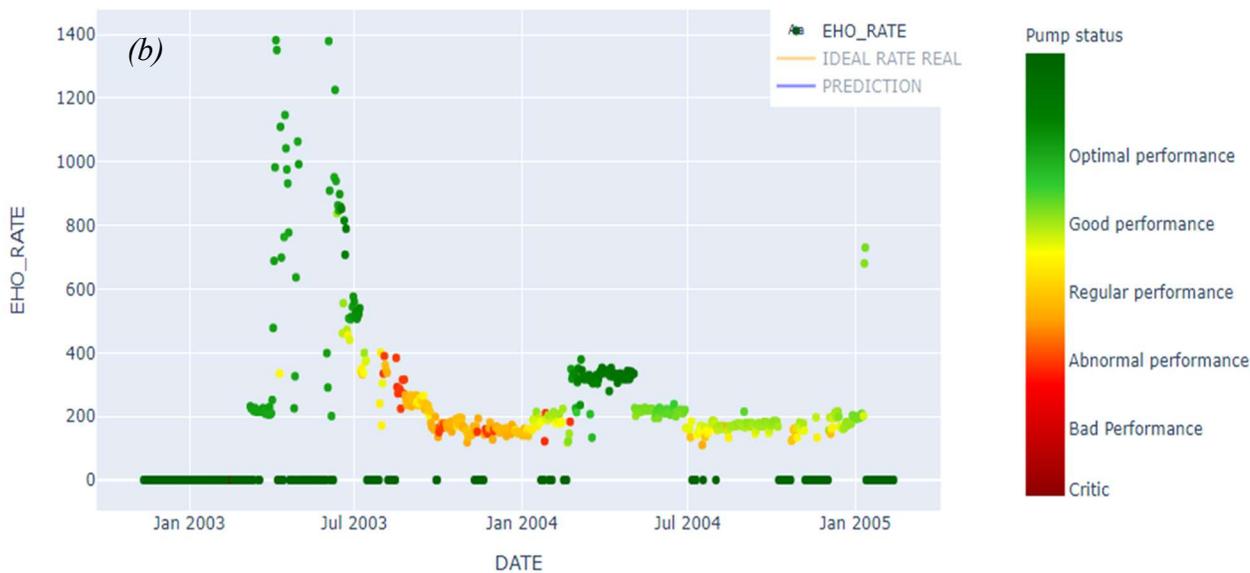


Figura 73: (a) Análisis del comportamiento de la bomba 1 en el pozo tipo respecto a la predicción
(b) Estado de la bomba 1 en pozo VD20

En el caso de la bomba en estudio, el algoritmo detectó un comportamiento irregular a partir de agosto de 2003 hasta finales de febrero de 2004, comportamiento que se mantuvo por aproximadamente seis meses y donde no se tomó ninguna medida preventiva que pudiese mejorar el funcionamiento de la bomba. Sin embargo, el software realizado genera una alerta de acuerdo con el caso caracterizado, como se muestra en la figura 74a y adicionalmente genera un grafica dinámica que permite comparar cuales parámetros han cambiado repentinamente respecto a la tasa para poder generar una solución rápida que permita volver a su funcionamiento original como se observa en la figura 74b.

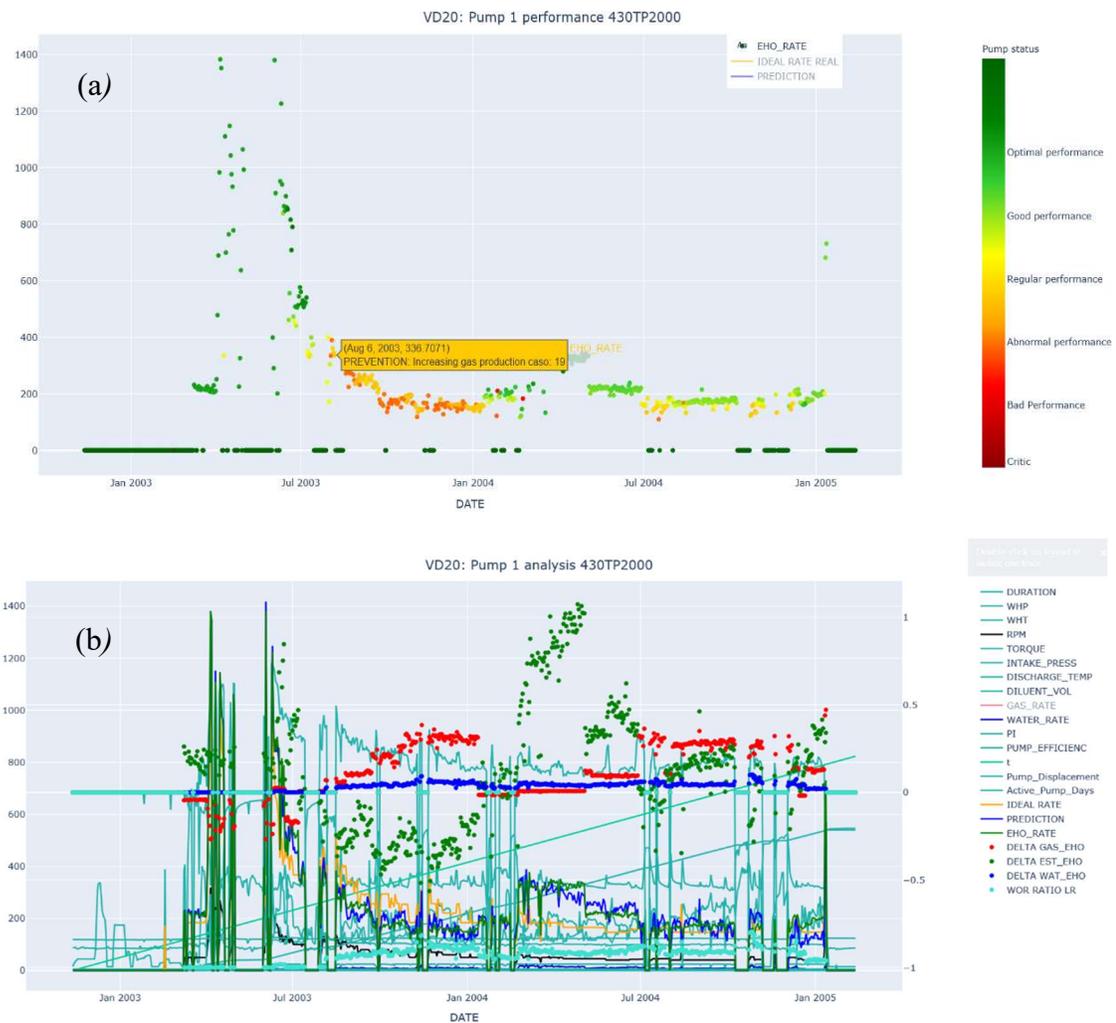


Figura 74: (a) Problema determinado por el algoritmo en la bomba 1, pozo VD20
(b) Parámetros de estudio en el pozo tipo

El análisis realizado a los distintos pozos permitió definir 6 causas principales que generan problemas de operatividad y pérdidas de producción en la macolla en estudio: aumentos excesivos de gas, cambios muy bruscos en el diferencial de presión, aumentos excesivos de la tasa de agua, cambios bruscos en el torque posible consecuencia de producción de sólidos, exceso de diluyente que genera un aumento del torque y aumentos excesivos de la producción de gas y por último fallas críticas por pérdida de fluidos.

4.4.1.- Aumentos excesivos de gas

El exceso de producción de gas es uno de los parámetros que más influye en la eficiencia de las bombas BCP, ya que disminuye de forma drástica la producción de petróleo. La figura 75a muestra el período de producción de la bomba 1 del pozo VD01, donde se muestra que, a partir de febrero de 2005, la bomba empieza a mostrar un comportamiento regular que se mantiene por aproximadamente 7 meses e inmediatamente el programa identifica una falla crítica por el exceso de producción de gas, sin embargo, en el periodo de comportamiento regular, el programa tiene la capacidad de anunciar cual es el parámetro que está afectando la producción. La figura 75b muestra el resultado de la gráfica dinámica que permite comparar todos los parámetros de producción en el período de estudio, en este caso el parámetro que afectó la eficiencia de la bomba es el gas, que como se muestra aumenta progresivamente.

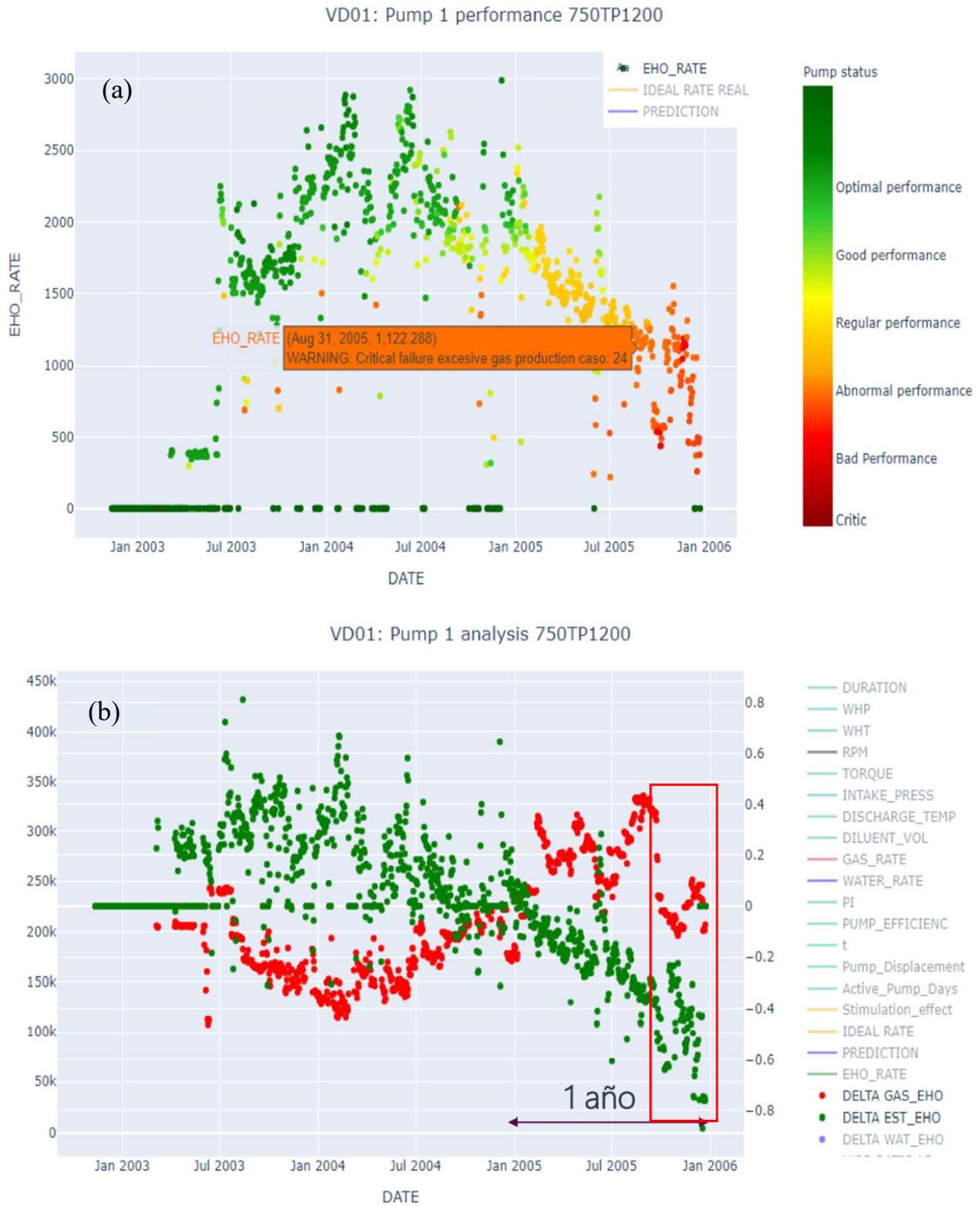


Figura 74: (a) Respuesta de la tasa de petróleo por un aumento de gas en el pozo
 (b) Análisis de parámetros de la bomba 1 Pozo VD20

El gas y las sustancias corrosivas que lo acompañan generan daños en los elastómeros de las BCP, por lo cual, si este tipo de comportamiento se mantiene por tiempos prolongados es muy probable que la bomba presente una falla por fuga de fluidos, como se representa en la gráfica con el recuadro de color rojo, donde la producción de petróleo disminuye drásticamente al igual que la producción de gas.

De acuerdo con el análisis realizado, este tipo de comportamiento puede ser causado por aumentos drásticos en el diferencial de presión o irregularidades en la inyección de diluentes, sin embargo, para este caso no se pudo determinar con exactitud la causa ya que la base de datos para la fecha no contiene registros de ninguno de estos parámetros.

4.4.2.- Cambios muy bruscos en el diferencial de presión

La presión de cabezal rige el comportamiento de las presiones dentro del sistema, por lo tanto, cualquier cambio que se realice en este se reflejará directamente en el comportamiento de la producción. La figura 76a muestra el comportamiento de la bomba 3 del pozo VD03, en este caso el pozo muestra una caída de producción a partir de enero de 2007, dicho comportamiento se mantiene por 9 meses aproximadamente. Cuando se evaluaron los parámetros, se observó que la presión de cabezal se disminuye aproximadamente 50 psi, este cambio generó un aumento del diferencial de presión en el pozo y un aumento en la producción de gas, como consecuencia una disminución en la producción como se muestra en la figura 76b. En este caso el algoritmo tiene la capacidad de determinar una anomalía que, si se hubiese atendido a tiempo, el pozo hubiese recuperado su tasa sin consecuencias mayores. Es importante destacar que muchas veces al aplicar diferenciales de presión tan altos, genera un aumento de la producción de gas en el pozo que puede ocasionar daños en la bomba y como consecuencia una fuga de fluidos en el sistema.

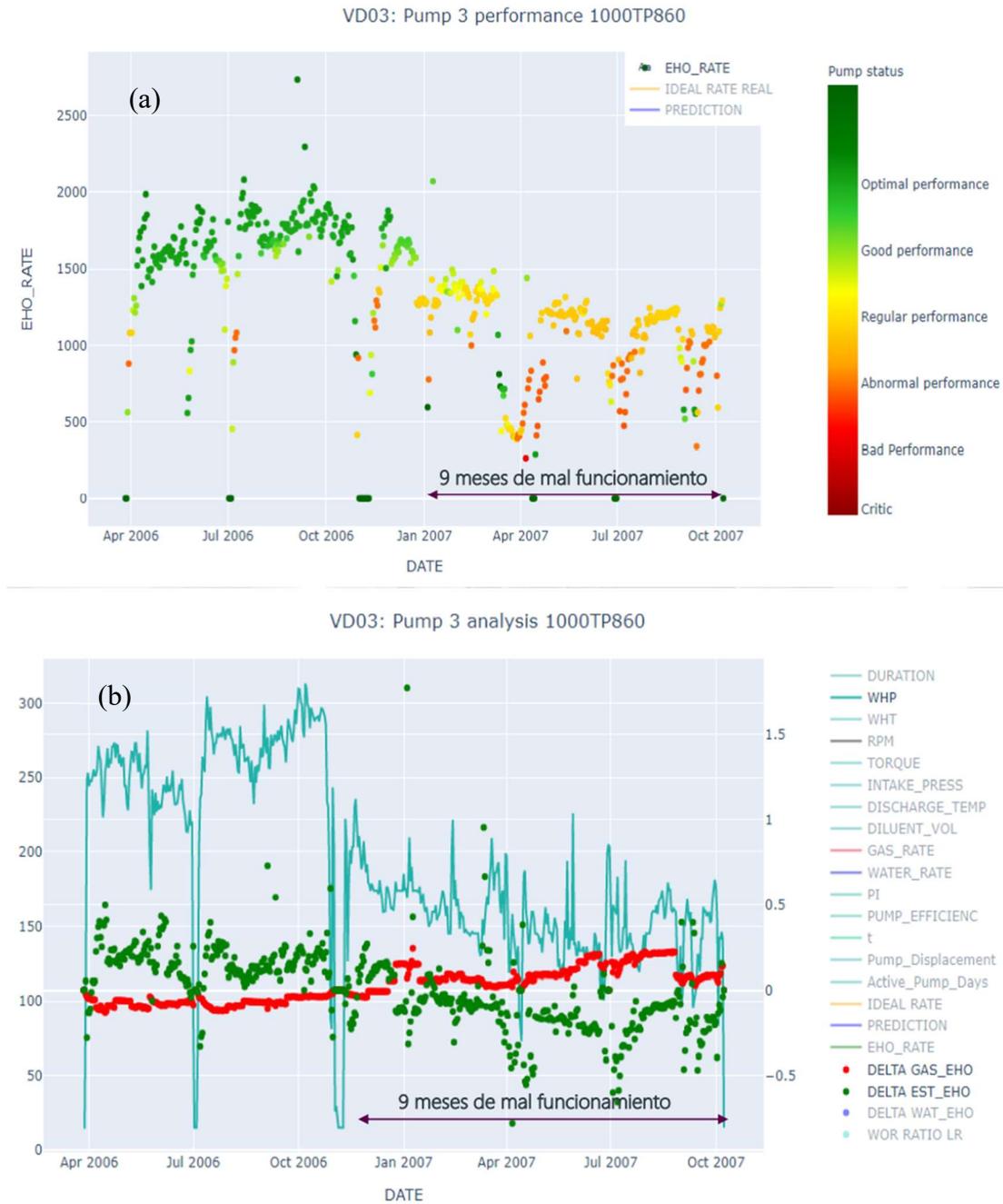


Figura 75: (a) Respuesta de la tasa de petróleo por cambio del diferencial de presión en el pozo VD03
 (b) Análisis del parámetro WHP de la bomba 3 Pozo VD03

4.4.3.- Aumentos excesivos de la tasa de agua

La zona de estudio presenta un acuífero asociado. En muchos casos debido a la alta permeabilidad vertical (movilidad del agua respecto al bitumen) el agua empieza a canalizarse dentro de los pozos. Este tipo de comportamientos se ve reflejado en el historial de producción, por lo tanto, el algoritmo tiene la capacidad de determinar este tipo de problemas y alertar al departamento de producción para que puedan tomar las medidas necesarias. En la figura 77a se muestra el resultado obtenido en el análisis del comportamiento de la bomba por efecto del agua, donde se observó que a partir del mes de abril de 2009 la producción de petróleo en el pozo cae un aproximado de 400 bpd de forma repentina. Este problema es anunciado por el software de forma inmediata, al realizar el análisis de los parámetros de producción se observó que la producción de agua respecto al petróleo incrementó considerablemente como se muestra en la figura 77b, donde los puntos de color azul oscuro representan la relación de producción de agua respecto al petróleo, los puntos de color verde representan la producción de petróleo, la producción de gas está representada en puntos de color rojo y por último la producción de agua respecto a la tasa total de líquido está representada en color turquesa. Al analizar el período de producción con anomalías en la figura 77a, se observó que el agua era el problema principal de la caída de producción, el cual se mantuvo constante por un periodo de aproximadamente dos meses, por lo tanto, el personal de mantenimiento y control de agua podían aplicar alguna medida preventiva o de control para evitar las pérdidas de producción. El resultado obtenido no correspondía a un problema en el funcionamiento de la bomba, y, por lo tanto, no era necesario realizar un cambio de ésta.

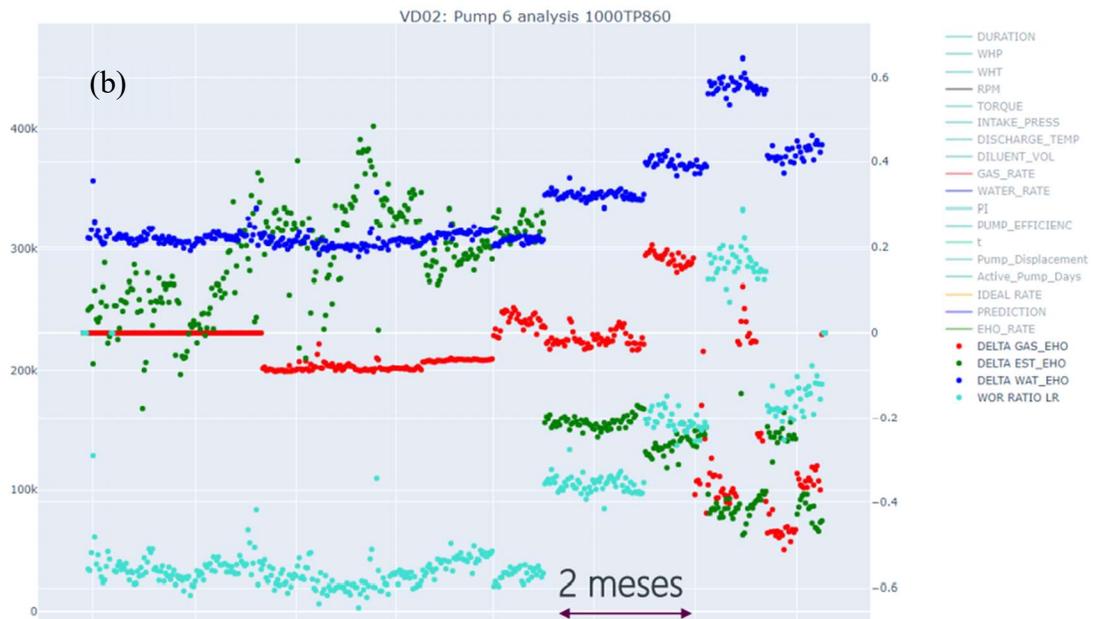


Figura 76: (a) Respuesta de la tasa de petróleo por producción de agua en el pozo VD02
 (b) Análisis del parámetro *Water_rate* y WOR de la bomba 6 Pozo VD02

4.4.4.- Cambios bruscos en el torque

Los cambios bruscos en el torque es otro de los problemas comunes al analizar el comportamiento de las bombas en la macolla. Este tipo de comportamiento se debe a la producción constante de sólidos dentro del pozo, genera un aumento en la fuerza de roce entre los equipos y por lo tanto una respuesta en el torque leído por los sensores, sin embargo, este tipo de comportamiento también puede visualizarse cuando hay cambios en el diluyente. El análisis realizado a este parámetro permitió definir el torque promedio de cada modelo de bomba a la profundidad de cada pozo, por lo tanto, este parámetro fue fundamental para referenciar y determinar anomalías asociadas. En la figura 78a se representa el resultado de un pozo con una caída de producción que se mantiene constante por aproximadamente 9 meses, sin embargo, el algoritmo determina que existe un comportamiento regular durante este periodo, respecto al funcionamiento de la bomba. Al analizar los parámetros de producción que presentaban cambios, se observó que el torque fue uno de los parámetros con mayores cambios. El estudio realizado al torque generado por el modelo de bomba 1000TP860 en estudio, mostró un torque promedio de 750 a 1050 lb/ft, mediciones por encima de este valor se consideran anómalas., por lo tanto, el comportamiento del torque que se muestra en la figura 78b representa un problema en el sistema.

El torque elevado fue un comportamiento que se presentó durante aproximadamente 4 meses manteniendo los rangos de producción, sin embargo, al transcurrir los 4 meses, el torque comenzó a disminuir nuevamente, pero, la producción del pozo disminuyó. Este tipo de comportamiento puede estar asociado a un posible taponamiento en el sistema pozo yacimiento, que generó una caída de la producción de petróleo, donde el torque se tiende a estabilizar ya que no arrastra tantos sólidos como antes al sistema. En la mayoría de los casos analizados, estos problemas no están relacionados con una falla en el funcionamiento de la bomba, sino, con problemas de obstrucción en el sistema que pudieron mejorarse o evitarse tomando en cuenta los 4 meses de comportamiento anómalos que presentó el torque, por lo tanto, en este caso

no era necesario realizar un cambio de bomba, sino estudiar si la bomba presentaba solidos en su interior o era un problema directamente de la arena.

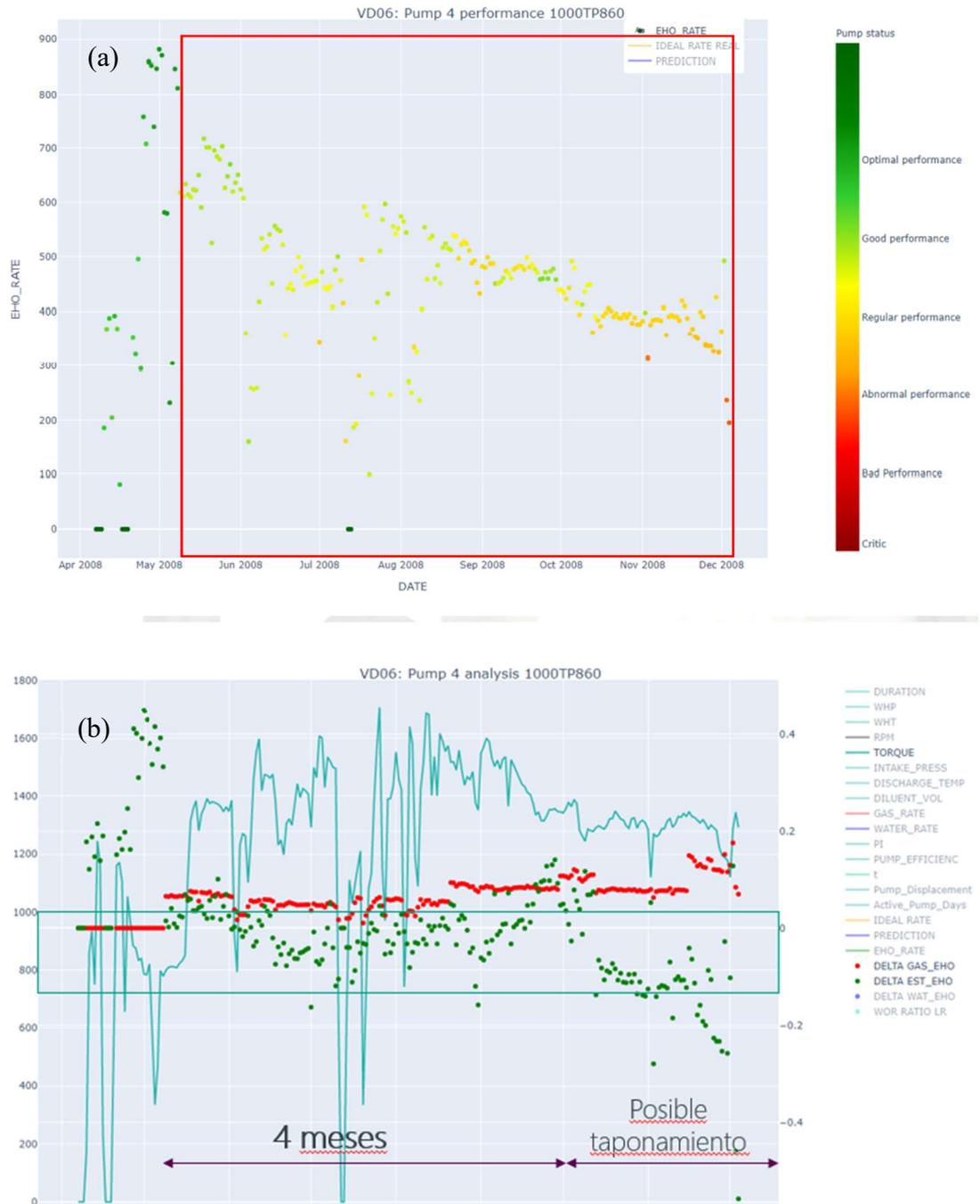


Figura 77: (a) Respuesta de la tasa de petróleo por cambios brucos en el torque posible producción de sólidos en el pozo VD06
 (b) Análisis del parámetro torque de la bomba 4 Pozo VD06

En la mayoría de los casos en estudio, los excesos de torque resultaban en una caída drástica de la producción que podría estar asociado a producción de arena en el sistema. El resultado de este comportamiento reflejó una gran cantidad de cambios de bombas innecesarios, lo cual se ve reflejado en una caída de la producción de petróleo a valores que con el tiempo se mantenían constantes, sin importar los cambios de bomba realizados. La figura 79 muestra el comportamiento de producción en un conjunto de 3 bombas consecutivas en el pozo VD35. La figura 79a muestra el comportamiento de producción de petróleo en la segunda bomba del pozo, como se aprecia en la figura, los datos de producción no pudieron caracterizarse de acuerdo a la escala de color, este problema se presentó en los casos en los que las bombas no aparecían en ninguna base de datos de las marcas registradas, por lo tanto, el algoritmo no puede determinar su desplazamiento nominal ni generar conclusiones acerca de su funcionamiento. Por otro lado, este periodo de producción es bastante corto y particular, desde el 22 de abril, la tasa de producción era de 420 bpd, sin embargo, 15 días después, exactamente para el 6 de mayo, la tasa de producción cayó a 280 bpd, una caída bastante drástica en el sistema. Al analizar los parámetros de producción y operatividad como se muestra en la figura 79b, se pudo observar que el torque aumentó considerablemente desde los primeros 5 días de producción. Éste comportamiento como se explicó con anterioridad representa una anomalía que puede estar asociada a la producción de sólidos que si no es atacada a tiempo puede generar un taponamiento en el sistema pozo yacimiento, como se presenta en este caso. La primera bomba que se muestra en la figura fue cambiada con tan solo 2 meses de funcionamiento y se colocó una nueva bomba modelo 750TP1200. En la figura 79c, se muestra el comportamiento de producción de la bomba, que, desde el inicio, el algoritmo determina que la bomba está trabajando adecuadamente bajo las condiciones actuales del sistema, sin embargo, la producción de petróleo está por debajo de la última tasa registrada en el periodo de bomba anterior. Al estudiar el comportamiento del torque, la tendencia se mantiene, dando como respuesta un torque elevado como se muestra en la figura 79d. Esta bomba empezó a funcionar el 21 de junio aproximadamente y para mediados de agosto, las consecuencias de un torque elevado se vieron reflejadas en la

caída de la producción de petróleo nuevamente a 190 bpd. Para el 23 de septiembre la bomba fue remplazada, éste cambio de bomba fue el segundo en un período menor a los 6 meses. Por último, la figura 79e, muestra el comportamiento de la nueva bomba modelo 750TP1200, donde se observó que el funcionamiento desde el comienzo de la producción está presentando fallas con una producción menor a la última registrada y un torque excesivo como se muestra en la figura 79f. Este comportamiento se mantuvo por un periodo aproximado de 6 meses, hasta que el operador del pozo decidió parar el funcionamiento y realizar un trabajo dentro del sistema pozo yacimiento restaurando la producción a los niveles expresados en la primera bomba (420 bpd) y disminuyendo el torque a los rangos normales. Sin embargo, no se pudo continuar con el análisis del comportamiento, ya que, la base de datos disponible contaba con datos hasta el año 2013.

Este estudio permitió concluir que en el campo se realizaron cambios de bombas innecesarios, que pudieron haberse evitado con solo aplicar las medidas correctivas a la producción de solidos a tiempo y adicionalmente obtener mayores tasas de petróleo durante este período.

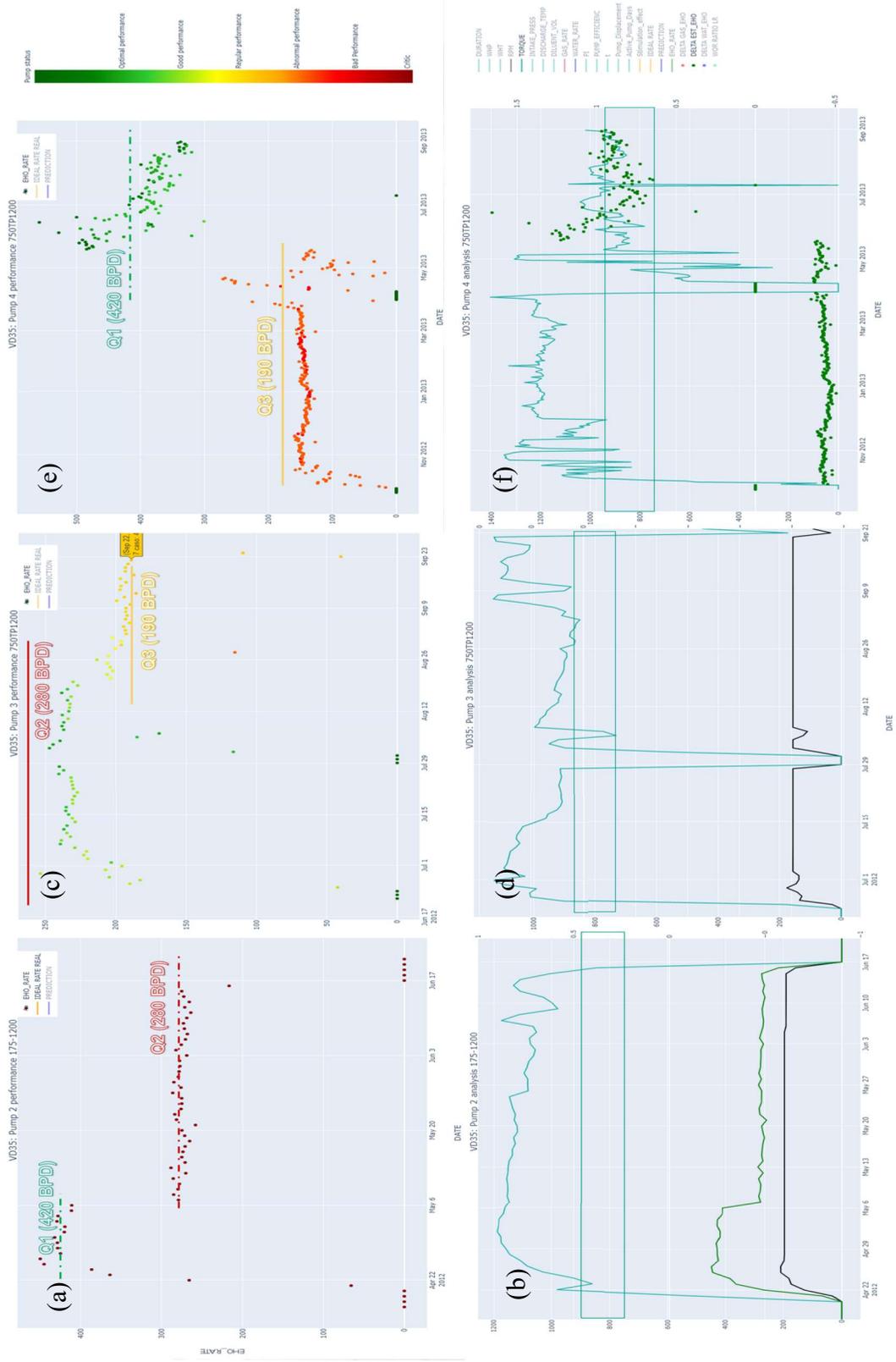


Figura 78: (a) Comportamiento de la producción de petróleo bomba 2 pozo VD35 (b) Análisis del torque de la bomba 2 pozo VD35 (c) Comportamiento de la producción de petróleo bomba 3 pozo VD35 (d) Análisis del torque de la bomba 3 pozo VD35 (e) Comportamiento de la producción de petróleo bomba 4 pozo VD35 (f) Análisis del torque de la bomba 4 pozo VD35

4.4.5.- Exceso de inyección de diluyente

En el estudio realizado se observó que existen gran cantidad de pozos con inyección de diluyente no controlada, por lo tanto, el exceso de diluyente en el sistema puede generar grandes pérdidas de producción. En la figura 80a se muestra el comportamiento de producción del pozo VD20, durante el periodo de funcionamiento de la tercera bomba, en el cual se observó que, para abril de 2006, el pozo presentó una caída de producción de aproximadamente 200 bpd. Este comportamiento se mantuvo por aproximadamente 6 meses hasta que se cambió la bomba, sin realizar ningún análisis para determinar con exactitud la falla. Al analizar la figura 80b se observó que la producción de petróleo disminuyó drásticamente debido a una inyección de diluyente excesiva. Ésta generó una disminución en la viscosidad del petróleo aumentando así la producción de gas en el sistema, por lo tanto, se determinó que el cambio de bomba era innecesario y que la caída de producción estaba asociada a las propiedades del fluido generadas como consecuencia de la inyección.

Este tipo de problemas puede ser muy perjudicial para la producción, por lo tanto, es necesario prestarles atención a los controladores de inyección y chequear constantemente los sensores que toman estas medidas.

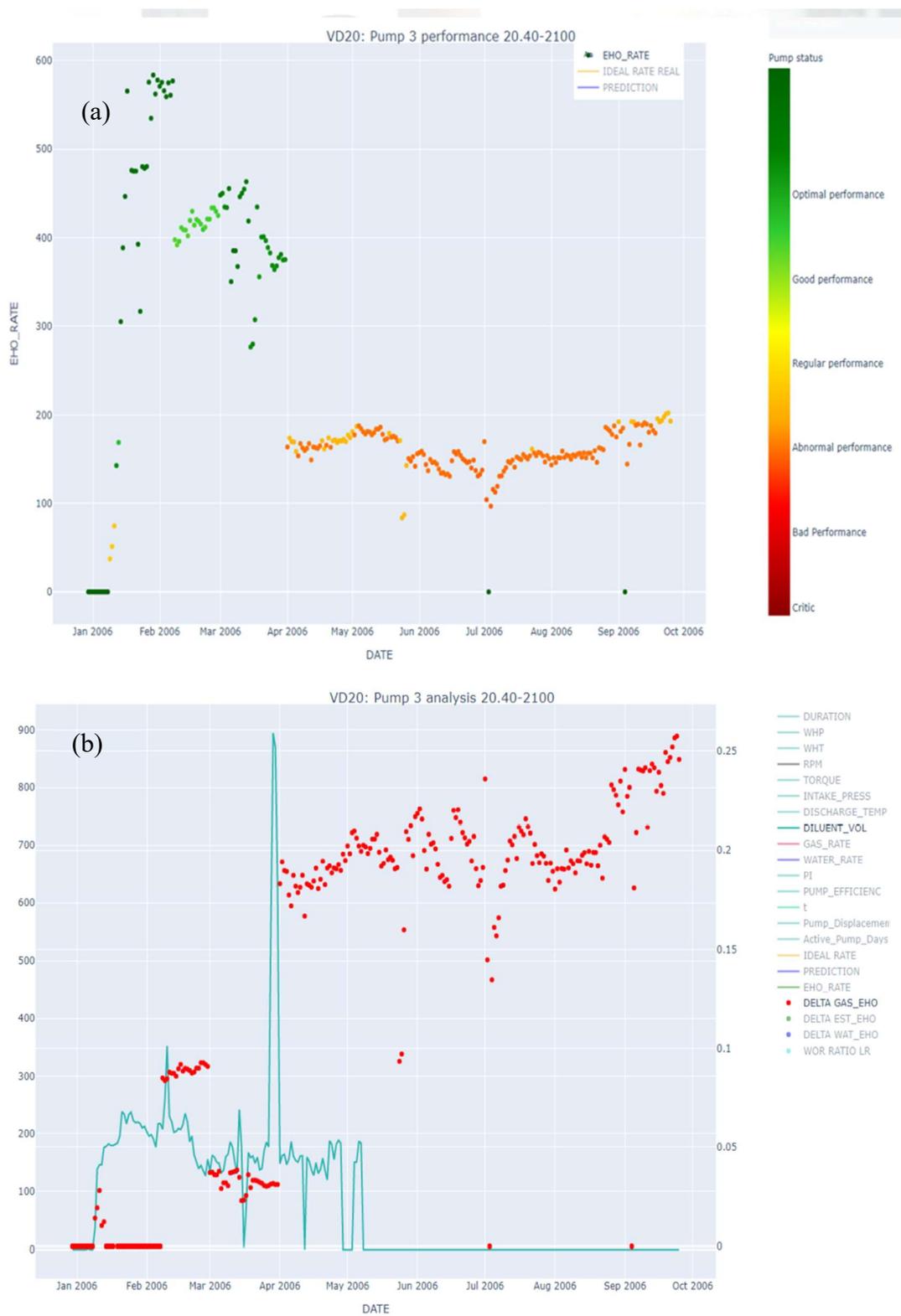


Figura 79: (a) Comportamiento de la producción de petróleo bomba 3 pozo VD20
 (b) Análisis de la inyección de diluentes en el período de la bomba 3 pozo VD20

4.4.6.- Fallas críticas por pérdida de fluidos

Este último problema está asociado directamente al funcionamiento de la bomba o problemas en la tubería de producción. Por lo tanto, el torque, al igual que, el comportamiento de todos los fluidos y las presiones, son parámetros fundamentales para determinar este tipo de fallas. En la figura 81a, se muestra el comportamiento de la bomba 1 del pozo VD18, donde se observó que, para mediados de septiembre de 2003, el pozo presentó una caída de producción de más de 1000 barriles por día. Al estudiar el comportamiento de los parámetros, en la figura 81b se observó que las tasas de todos los fluidos presentaban una caída drástica y adicionalmente el torque disminuyó considerablemente respecto a sus valores promedios, este tipo de comportamiento representa una fuga de fluidos en el sistema que requiere de atención inmediata, en esta caso tras una evaluación en el mes de julio de 2003, y diversas pruebas en los meses posteriores se determinó que la falla estaba asociada a la bomba, y, por lo tanto se procedió a reemplazarla en noviembre de 2003.

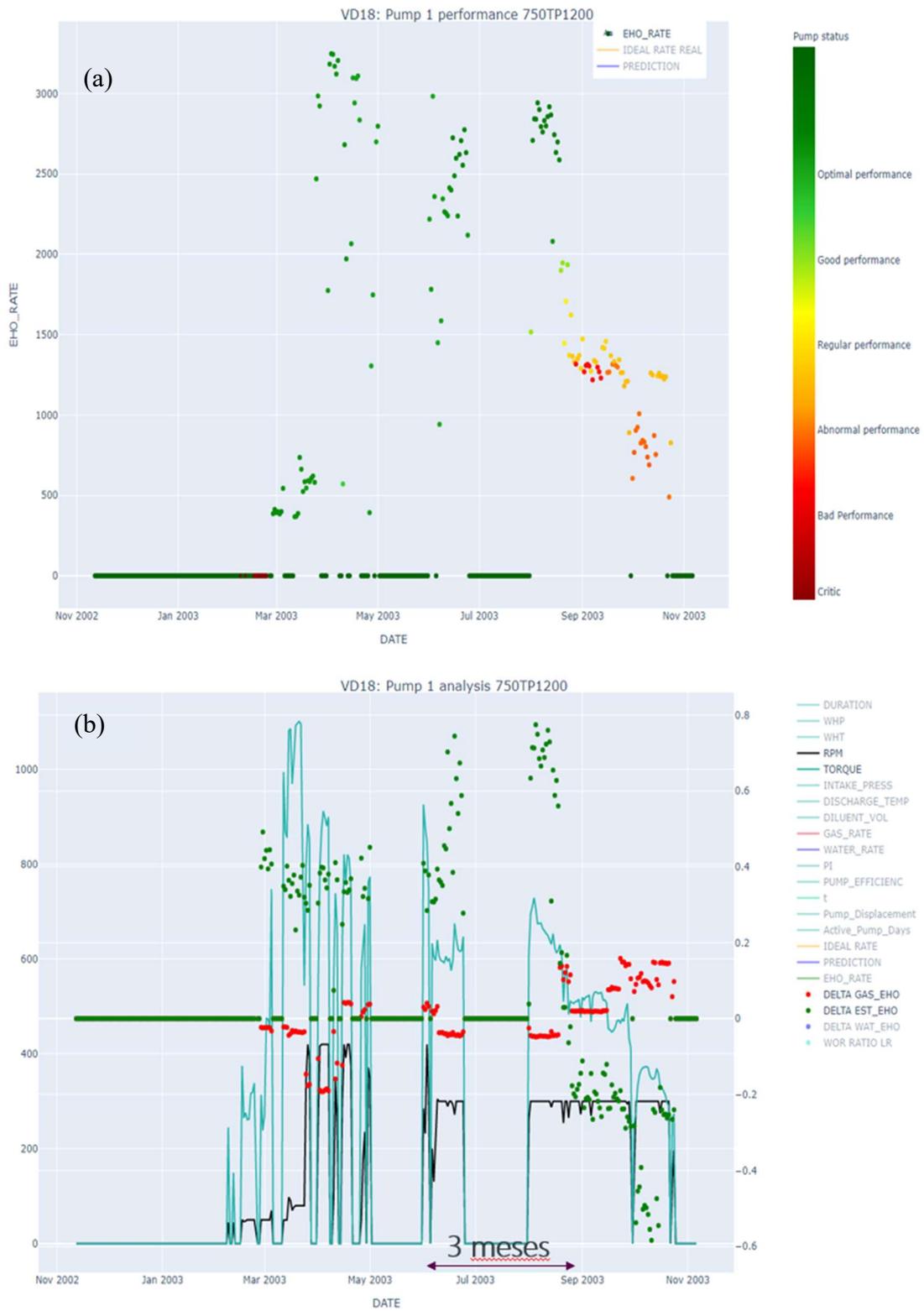


Figura 80: (a) Comportamiento de la producción de petróleo bomba 1 pozo VD18
(b) Análisis del torque y las tasas de producción en el período de la bomba 1 pozo VD18

Como se pudo observar, ésta herramienta te permite determinar cuando la bomba presenta un mal funcionamiento, con la finalidad de realizar un análisis de los parámetros que puedan estar afectando y poder tomar decisiones optimizando tiempo y recursos.

4.4.7.- Análisis final de las bombas en la macolla en estudio

El análisis realizado a la macolla permitió determinar de forma general la cantidad de cambios de bombas necesarios e innecesarios y estudiar la eficiencia de las bombas en el campo, a través de la significancia estadística, los días de producción y el porcentaje de barriles producidos respecto a sus valores nominales.

En el estudio realizado se lograron analizar un total de 261 bombas, de las cuales 182 cambios de bomba se consideraron como innecesarios (69,72%), 44 cambios de bombas necesarios (16,86%) y 35 cambios de bombas no concluyentes (13,42%).

De los cambios de bombas innecesarios, los cuales se muestran en la figura 82, 124 fueron periodos de tiempo que no tuvieron control de diluyente, 114 con exceso de torque, 98 con posible producción de sólidos, 86 con posible taponamiento, 34 con exceso de agua, 18 con exceso de gas y 3 pozos inundados.

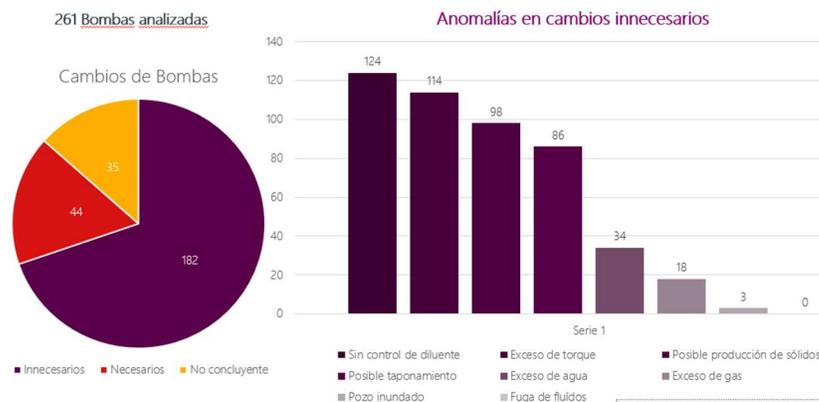


Figura 81: Análisis de cambios de bombas innecesarios

Por otro lado, se pudo determinar que, de los 44 cambios de bombas necesarios, 31 cambios de bombas se debieron a un exceso de diluyente, que produjo daños en el elastómero o exceso de la producción de gas, el cual contiene agentes corrosivos que dañaron el sistema de la bomba. 28 casos presentaron fuga de fluidos, 20 casos posible producción de sólidos, 5 casos exceso de gas, 4 casos con posible taponamiento, 3 casos con pozo inundado (con una alta relación agua petróleo) y 2 con exceso de agua como se muestran en la figura 83.

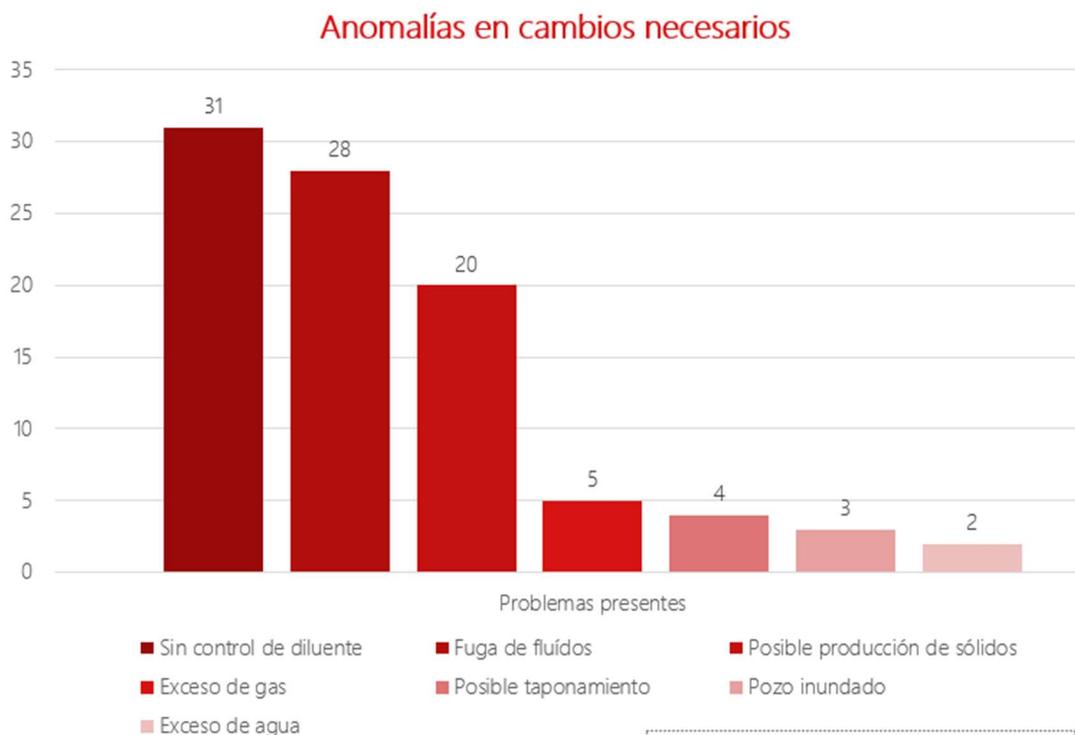


Figura 82: Análisis de cambios de bombas necesarios

Es importante resaltar que los casos mencionados en ambos análisis no son excluyentes, por lo tanto, un pozo puede presentar 2 o más anomalías.

Al analizar los cambios realizados de acuerdo con el modelo y marca de la bomba, se obtuvieron los resultados que se muestran en la figura 84. El modelo más usado es el 1000TP860 que corresponde a la marca PCM MOINEAU, de los cuales

60 cambios de bomba fueron innecesarios (73,17%), 19 cambios necesarios (23,17%) y 6 no concluyentes (3.66%). El resto de los modelos de las bombas se usaron en menos de 30 oportunidades, sin embargo, los resultados obtenidos de los otros modelos se pueden observar en la gráfica.

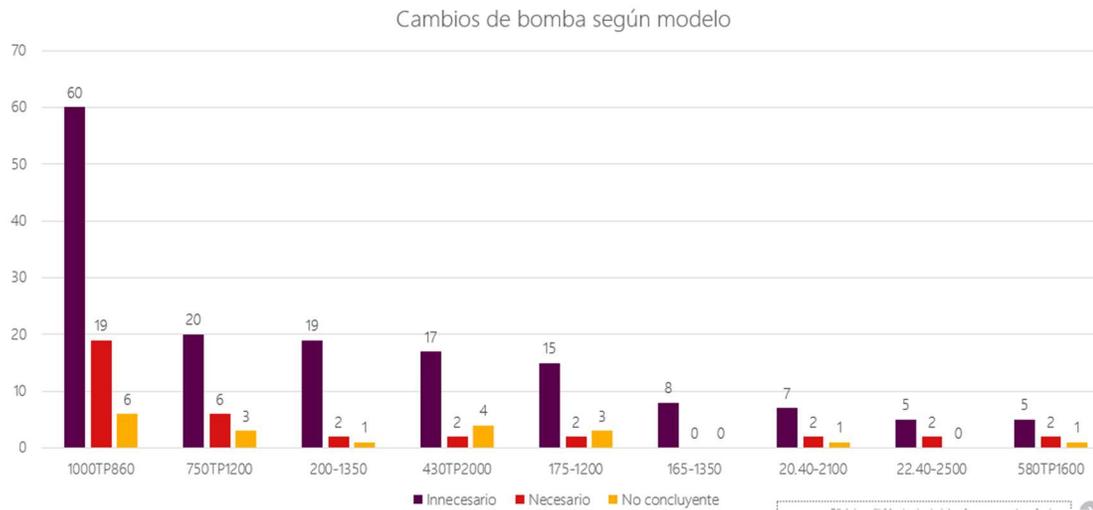


Figura 83: Análisis de cambios de bombas de acuerdo con el modelo

Por otra parte, se analizó la duración de cada bomba y la productividad de estas respecto al modelo como se muestra en la figura 85, donde se observó que la bomba con mayor duración era el modelo 750TP1200 de la marca PCM MOINEAU con una duración promedio de 614 días y un porcentaje de real de producción de 42,18%. Sin embargo, la bomba con mayor porcentaje real de producción fue la 430TP2000 (56.01%), pero con una duración menor de 340 días. Es importante destacar que la duración de las bombas es un valor con mucha incertidumbre, ya que, este valor depende de los cambios de bombas realizados y durante la investigación se demostró que gran cantidad de cambios de bombas fueron innecesarios.

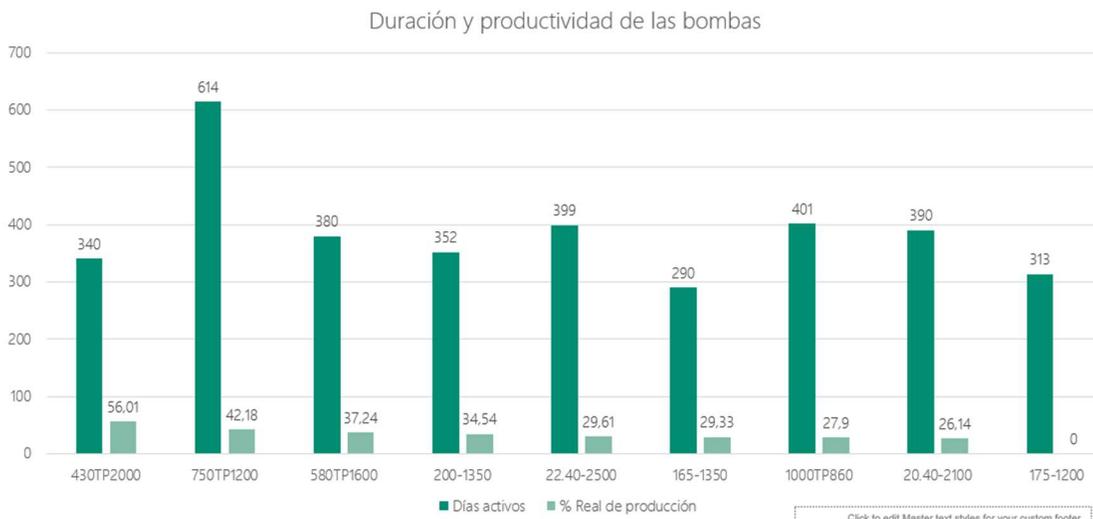


Figura 84: Análisis del comportamiento de los modelos de bombas respecto a su duración y productividad.

Por otra parte, al estudiar la significancia estadística de cada uno de los modelos de las bombas y su desplazamiento, se pudo observar que estos parámetros varían de acuerdo con el pozo en estudio (Anexo 9). Sin embargo, se pudo observar que entre menor era la significancia estadística del modelo de la bomba, el rendimiento de esta respecto a la producción de petróleo se mantenía por mayor cantidad de tiempo con un buen funcionamiento.

En la siguiente sección se realiza un análisis completo para comparar el modelo realizado, respecto al modelo actual de seguimiento.

4.5.- Validación de la precisión del modelo realizado con el modelo actual de seguimiento

Con la finalidad de cumplir el cuarto objetivo de la investigación fue necesario analizar y comparar ambos modelos.

El modelo desarrollado en este TEG fue realizado con datos de producción y operación desde el año 2002 al año 2013, por lo tanto, se buscó comparar el modelo y

los resultados aquí obtenidos con el modelo de seguimiento para el periodo 2014 al 2018.

Para la validación de este tipo de metodologías es necesario contar con la misma cantidad de columnas de datos (parámetros) y en el mismo formato de tiempo con los cuales fue realizado el entrenamiento del modelo inicial, si no, el algoritmo no podrá realizar una predicción precisa.

El primer problema que se presentó en la base de datos 2014 -2018, fue el formato de tiempo, donde los datos por pozo se reportan de forma mensual, mientras que, en la base de datos 2002-2013 se reportan diariamente. Adicionalmente, la cantidad de parámetros de entrada es menor en el periodo 2014 - 2018. Esta incongruencia entre los datos no permitió que la validación cuantitativa del modelo original fuese posible, sin embargo, se creó un nuevo modelo con los datos promediados de manera mensual y basado solo en los parámetros de entrada disponibles en la nueva versión, dando como resultado un nuevo modelo con mayor error e incertidumbre, debido a que la cantidad de datos por columna fue menor y el número de filas en estudio disminuyó de 365 a tan solo 12 filas por año. Por esta misma razón, el análisis del comportamiento de las bombas no fue posible.

En la figura 86a se muestra el resultado del modelo lineal por mes en el periodo 2002-2013, donde el error entre la predicción y el modelo es bajo, y el modelo es capaz de reproducir la tendencia entre los picos en los periodos de prueba. Sin embargo, en la figura 86b, se muestra el gráfico resultante del modelo en el periodo 2014-2018, donde las tasas de producción se presentan como líneas rectas por periodos prolongados y adicionalmente, la producción reportada indica que los pozos trabajaron continuamente por un mes a altos valores de RPM, con una producción mensual de 6 bpd, sin reportar valores congruentes en la tasa de producción de agua o gas.

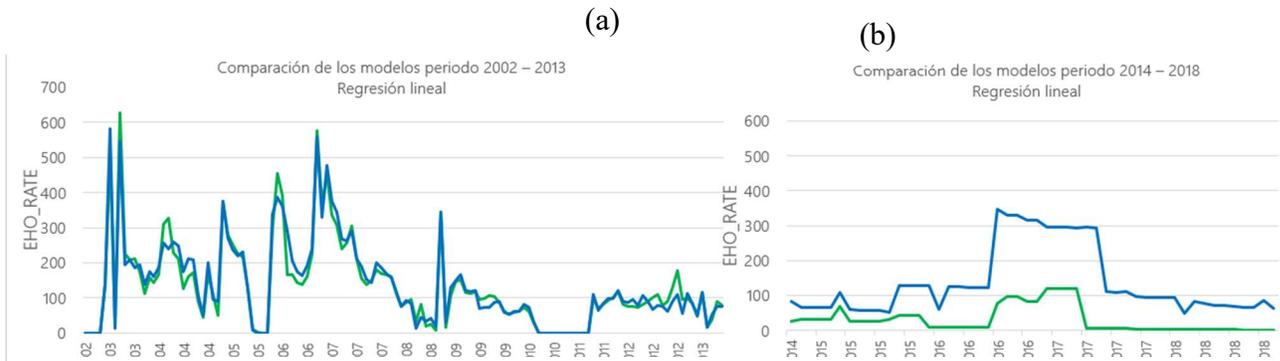


Figura 85: (a) Comportamiento del modelo de predicción en el año 2002 -2013
 (b) Comportamiento del modelo de predicción en el año 2014 -2018.

La disminución de monitoreo de datos diarios a mensuales realizado a partir del año 2014 fue decisiva en cuanto a la imposibilidad de aplicar este estudio. Adicionalmente, se observa que los datos obtenidos fueron manipulados a fin de obtener un valor mensual promedio, y se desconoce si dichos datos fueron filtrados, verificados y procesados antes de promediarlos, aumentando así la incertidumbre del modelo.

A pesar de que el modelo no pudo validarse de forma cuantitativa, el análisis realizado al conjunto de datos permitió validar el modelo de forma cualitativa, donde se puede observar que la calidad y la cantidad de los datos es fundamental para poder validar la precisión de los modelos. Por otro lado, los tiempos de respuesta y análisis de los datos recibidos se puede realizar en fracción de segundos y determinar si los mismos tienen sentido físico. A pesar de que se cuenten con sistemas automatizados de medición, es necesario verificar si los datos cargados en el sistema son correctos y congruentes. Por otra parte, la adquisición de datos mensual no permite un análisis detallado de lo que ocurre en el campo, ni logra un análisis robusto del comportamiento de las bombas que permita detectar fallas o anomalías a tiempo.

4.6.- Etapa de despliegue

El despliegue es la última etapa de la metodología CRISP-DM seleccionada, la cual permitió cumplir con el objetivo general de la investigación. El resultado del despliegue es una interfaz gráfica de fácil acceso y manejo para los usuarios.

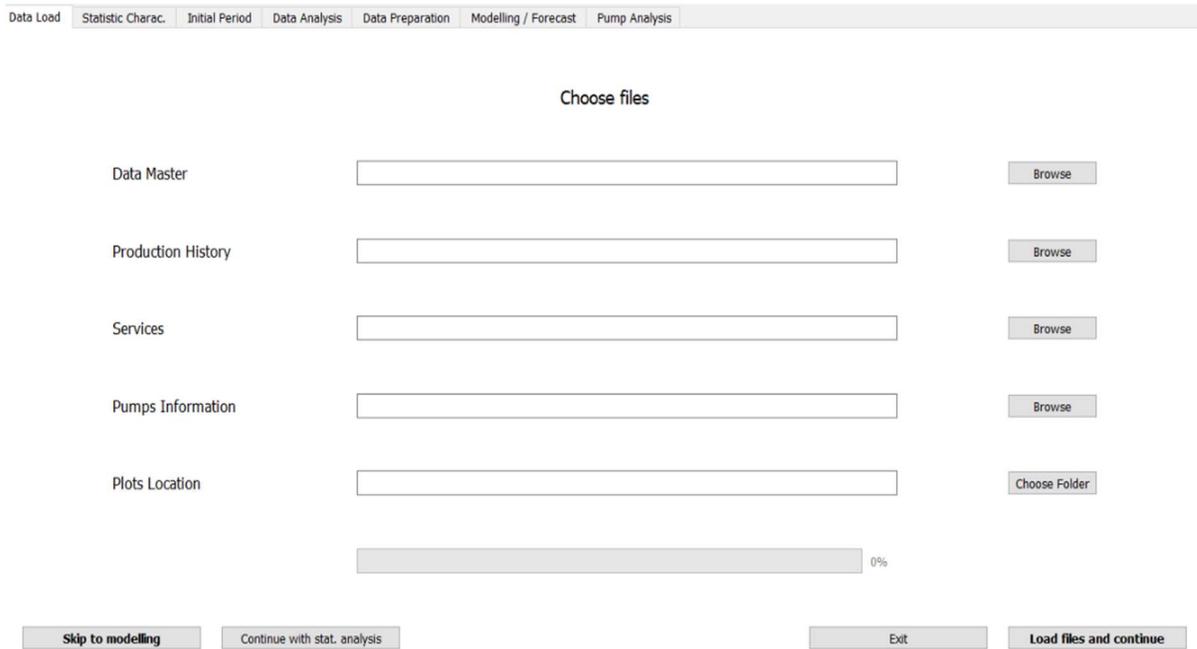


Figura 86: Panel inicial de la interfaz gráfica

En la figura 87, se observa la pantalla inicial del proceso el cual comprende la carga de datos. En las pestañas superiores se observan los distintos procesos como: la caracterización estadística, el estudio del periodo inicial, el análisis de datos, la preparación de datos, el modelado y predicción y por último el análisis de las bombas.

Todas estas pestañas fueron configuradas con un set de botones que permiten generar la interconexión entre cada uno de los procesos antes mencionados y generar los archivos y gráficos dinámicos que pueden visualizarse a través de cualquier navegador web.

El despliegue de la herramienta es necesario para que los usuarios puedan realizar el análisis de datos, por esta razón, se generó un manual (Anexo 10) para el usuario con la finalidad de guiar a las personas durante el uso de la aplicación. La

herramienta y los algoritmos desarrollados solo pueden ser modificados por los programadores.

En los últimos dos capítulos se explica todo el desarrollo de la metodología CRISP-DM aplicada a un conjunto de pozos, tomando en cuenta datos de producción y operatividad. Durante el análisis de resultado se logró comprobar que el uso de estas metodologías es eficiente y permite un rápido análisis de todos los parámetros involucrados.

CAPÍTULO V

CONCLUSIONES

A continuación, se presentan las conclusiones que demuestran el cumplimiento de todos los objetivos de la investigación. Este párrafo más bien comenzando las conclusiones.

- La calidad de las bases de datos influye de forma directa en la incertidumbre de los modelos generados. El procedimiento de caracterización estadística es eficiente y permite disminuir los errores en la etapa del modelado un 13,05% en promedio.
- El modelo de predicción que mejor se ajustó a los datos de producción en el campo en estudio es el modelo polinomial de grado 2 con un error relativo promedio de 10,65% y un error absoluto de 40 bpd por día. Mientras que para el análisis de las bombas la técnica que se ajustó a los objetivos de la investigación fue el árbol de decisiones con 24 casos.
- Las herramientas de *machine learning* y *deep learning* permiten obtener resultados aceptables en un muy corto periodo de tiempo, una vez que los datos están estructurados y los algoritmos definidos. La respuesta de estos algoritmos supervisados se basa en la experiencia de los usuarios y en la tendencia y

comportamiento de los parámetros, lo que permite generar alertas instantáneas cuando se presentan anomalías, facilitando así, la toma de decisiones.

- La significancia estadística es un indicador fundamental para determinar la influencia de un parámetro o característica en el atributo de salida, lo que le permite al usuario excluir variables, facilitando el análisis.
- El desarrollo de esta metodología permitió analizar exhaustivamente el comportamiento de las bombas, determinando si era necesario realizar un reemplazo de las mismas, o, si por el contrario, se hubiese extendido su periodo de vida útil con un mantenimiento preventivo. Se analizaron un total de 261 cambios de bombas, de las cuales, 182 fueron innecesarios, 44 necesarios y 35 no concluyentes. En el análisis de los cambios de bomba innecesarios se determinó que la causa de la caída de producción estaba relacionada con malas operaciones en el campo y falta de control en los parámetros operativos como x y Y .
- La metodología de ciencia de datos CRISP-DM es eficiente para el análisis de parámetros de operación y productividad de un pozo petrolero, ya que permite analizar el historial de producción con detalle, predecir la producción con un error relativo y absoluto aceptable (10.65% y 40 bbl respectivamente) y detectar anomalías en los sistemas de levantamiento de forma rápida, cumpliendo con los objetivos planteados en la investigación.
- No es posible realizar la validación cuantitativa del modelo, ya que no se cuenta con una base de datos diaria y con todos los parámetros estudiados, sin embargo, se logró una validación cualitativa, determinando que la rapidez de respuesta y análisis con este tipo de modelos permite identificar tanto fallas en los sistemas operacionales, como problemas en los sistemas de adquisición de datos de forma casi inmediata.
- Los procesos de simulación numéricas requieren de una considerable cantidad de tiempo de procesamiento ya que utilizan parámetros muy rigurosos y una gran cantidad de datos. Mientras que los algoritmos de aprendizaje automatizado pueden obtener resultados similares en menor tiempo y no

requiere parámetros tan rigurosos como los procesos de simulación deterministas.

- La herramienta computacional generada contiene un aproximado de 206 algoritmos desarrollados, con un total de 13290 líneas de código que permiten la implementación de todas las etapas de la ciencia de datos, herramientas de *machine learning* y *deep learning* por medio de una sola interfaz gráfica.
- Se destaca que la metodología y el software desarrollado en este TEG no es exclusivo para el campo en estudio, sino que puede aplicarse a diferentes áreas de la industria y a distintos campos.

RECOMENDACIONES

- Realizar un seguimiento exhaustivo a los datos recolectados del sistema automatizado durante las operaciones diarias basados en el formato del año 2002 – 2013, con la finalidad de garantizar la calidad de los datos.
- Estandarizar la información suministrada por cada departamento para futuras mejoras del software.
- Centralizar la recolección de datos para un manejo óptimo de la información, que permita intersectar todas las bases de datos y mejorar el análisis de los parámetros de producción y operatividad.
- Aplicar esta metodología a otras macollas del campo para mejorar el algoritmo y verificar el comportamiento de los pozos con el historial de producción actualizado.
- Incluir en el estudio las mediciones eléctricas del campo para mejorar el proceso de caracterización de fallas.
- Agregar al algoritmo las propiedades del yacimiento, distribución de los pozos en la arena, propiedades de los fluidos, potencial del pozo y estudios de la producción de arena para generar un análisis más robusto y que pueda servir de referencia para futuros estudios.
- Generar un sistema de alertas dentro del algoritmo que le permita al operador en campo conocer fallas en los sensores y capacitar al personal para que pueda caracterizar los datos anómalos, con la finalidad de no acumular grandes cantidades de datos que generen incertidumbre en los modelos.
- Es recomendable trabajar este tipo de algoritmos con computadoras de alta gama que permitan obtener respuestas rápidas en poco tiempo y sin ninguna limitación computacional (preferiblemente, equipos con GPU).
- Mantener actualizado el software con la finalidad de detectar, analizar y caracterizar nuevos comportamientos anómalos durante la operatividad de los pozos.

ANEXOS

Anexo 1: Acrónimos

A 1 Anexo 1 Lista de acrónimos

Acrónimo	Column1
API	American Petroleum Institute
lb/mol	Libras sobre mol
PCN	Pies cúbicos normales
BN	Barriles normales
BY	Barriles de yacimiento
cp	Centipoise
%AyS	Porcentaje de agua y sedimentos
Pe	Presión estática
Pwf	Presión de fondo fluyente
Pwh	Presión de cabezal
Ql	Tasa de líquido
Qw	Tasa de agua
Qo	Tasa de petróleo
RAP	Relación agua petróleo
RGP	Relación gas petróleo
IP	Índice de productividad
H	Sumergencia
PB	Distancia vertical de succión de la bomba
NE	Nivel estático
ND	Nivel dinámico
RPM	Revoluciones por minuto
H ₂ S	Ácido sulfúrico
BCP	Bomba de cavidades progresivas
bbbl/día	Barriles por día
m ³ /día	metros cúbicos por día
mD	milidarcy
FPO	Faja Petrolífera del Orinoco
WHT	Temperatura de cabezal
WHP	Presión de cabezal
SEMMA	<i>Sample, Explore, Modify, Model, Assess)</i>
CRISP-DM	<i>Cross-Industry Standard Process for Data Mining</i>
KDD	<i>Knowledge Discovery in Databases</i>
AI	<i>Artificial Intelligence</i>
ML	<i>Machine learning</i>
DL	<i>Deep Learning</i>
KNN	<i>K-nearest neighbors</i>

A 1: Lista de acrónimos (Continuación)

SVM	<i>Support Vector Machine</i>
ANN	<i>Artificial Neural Network</i>
RNA	<i>Red neuronal artificial</i>
TEG	<i>Trabajo Especial de Grado</i>
PCP	<i>Progressive Cavity Pump (bomba de cavidades progresivas, por sus siglas en ingles)</i>
BP	<i>British Petroleum</i>
IBM	<i>International Business Machines Corporation</i>
PVT	<i>Presion, Volumen y Temperatura</i>
FDA	<i>Functional Data Analysis</i>
SAS	<i>Statistical Analysis System</i>
MSE	<i>Mean Squared Error</i>
MAE	<i>Mean Absolute Error</i>
SPE	<i>Society of Petroleum engineers</i>
°F	<i>grado Fahrenheit</i>
d	<i>dia</i>
SCF	<i>Standard Cubic Feet (pies cubicos estandar por sus siglas en ingles)</i>
WOR	<i>Water oil Ratio (relacion agua petroleo por sus siglas en ingles)</i>
GOR	<i>Gas Oil Ratio (relacion gas petroleo por sus siglas en ingles)</i>
NaN	<i>Not a Number</i>
N/A	<i>Not Available</i>
EHO	<i>Extra Heavy Oil (petroleo extra pesado por sus siglas en ingles)</i>
ReLU	<i>Rectified Linear Unit</i>
ADAM	<i>Adaptive Moment Estimation</i>
RAM	<i>Random Access Memory</i>
GPU	<i>Graphics Processing Unit</i>

Anexo 2: Procedimiento manual para la optimización de un pozo con sistema de bombeo BCP del Ing. Nelvy Chacin

- A. Conociendo la curva de afluencia del pozo, se establece el nuevo caudal y, por ende, el nuevo nivel de fluido que se desea alcanzar. Este caudal debe ser igual o menor que el establecido por el departamento de yacimientos como caudal máximo. El nivel de fluido debe proveer una sumergencia mínima a la bomba de 200 – 300 pies.*
- B. Con este caudal se calculan las pérdidas por fricción en el reductor y en la línea de producción (se recomienda, si es posible medir esta última variable). Así mismo, se calcula la presión en la succión de la bomba, considerando el nivel de fluido y la presión de revestidor.*
- C. Se procede a estimar el diferencial de presión en la bomba sumando las pérdidas de presión totales con la presión de cabezal al resultado se le resta la presión de succión (ΔP en la bomba = Presión Descarga - Presión Succión) tal y como se expuso en el punto correspondiente a Diseño del Sistema en este documento; se verifica si el diferencial de presión calculado es igual o menor que la capacidad de la bomba, no es aconsejable someterla al máximo HEAD o diferencial de presión, pues disminuiría considerablemente la eficiencia y la vida útil del equipo, para ello es buena práctica no superar el 80% del máximo diferencial de presión (según catalogo). Si el diferencial de presión calculado es mayor que el 80% del máximo diferencial de presión de la bomba, se supone un caudal menor y se inicia el proceso (si el diseño fue acertado esta condición no debería presentarse).*
- D. Conociendo el caudal y el diferencial de presión, utilizando la curva característica de la bomba instalada se lee en la misma la nueva velocidad de operación.*
- E. Si la velocidad leída es mayor que 300 R.P.M. (lo cual no debería ocurrir ya que en la fase de diseño se utilizaron los factores de servicio correspondientes), se debe suponer una velocidad inferior y leer en la curva característica el caudal a producir a esta velocidad e iniciar el proceso (punto “A”), de lo contrario, seguir adelante.*
- F. Se calcula la potencia hidráulica y con ella el torque; también se calcula la tensión en la primera cabilla (la conectada al eje del cabezal rotatorio o a la barra pulida, según sea el caso), tomando en cuenta la fuerza neta que se genera entre la succión y descarga de la bomba por efecto de las presiones correspondientes.*
- G. Con el torque y la tensión de la primera cabilla se procede a calcular el esfuerzo combinado, el cual se compara con la tensión de fluencia, si este valor es mayor que un 80% de la tensión de fluencia, se supone un caudal menor y se inicia el proceso, en los casos de pozos desviados u horizontales el cálculo de la tensión es muy complicado y para ello se aconseja el uso de programas especializados.*

H. Por último, se procede a verificar si los equipos de superficie están capacitados para manejar las nuevas condiciones. La tensión de la primera cabilla se compara con la capacidad del cabezal de rotación, El torque con el torque disponible en la salida de la caja reductora y la potencia requerida no debe exceder el 80% de la potencia instalada. Si todas las condiciones resultan favorables se realiza el ajuste de velocidad.”

Finalmente, cuando las condiciones del pozo estén estables, se procede a elevar la cantidad de revoluciones por minuto de la bomba y se repite el procedimiento, hasta alcanzar la tasa máxima estimada para el pozo. Cuando esto ocurra, se puede decir que el pozo esta optimizado (entendiendo con esto que la relación beneficio/costo sea la mayor posible dentro del marco legal, y sin comprometer los limites seguros de operación de los equipos).^[13]

Anexo 3: Problemas típicos en BCP y posibles soluciones Hirschfeldt (2008),

De acuerdo con el Manual de bombeo de cavidades progresivas de Marcelo Hirschfeldt (2008), los problemas típicos en bombas BCP se pueden agrupar en distintos casos de acuerdo con los problemas que presente el caudal y el funcionamiento general de la bomba como los elementos que la conforman. Todos estos casos se resumen desde la tabla 29 hasta la tabla 39 que se presentan a continuación:

Anexo 3 Problemas típicos en BCP (Caso 1)^[24]

Caso 1: Bajo caudal y baja eficiencia volumétrica. La velocidad es la prefijada. El rango de corriente está normal	
Causa probable	Acción recomendada
El rotor no está totalmente insertado	* Verificar el espaciado y corregir de ser necesario.
La presión de descarga de la bomba es inferior a la necesaria	* Verificar la altura de elevación necesaria por cálculo. * Cambiar la bomba si es necesario.
El rotor está sub-dimensionado para la temperatura del pozo	* Revisar la temperatura y el tipo de rotor usado. * Cambiar el rotor de ser necesario.
Existe pérdida en la tubería	* Buscar el tubing roto y cambiar la unión.
Existe un alto RGP	* Proveer medios para las anclas de gas natural, instalando la bomba por debajo del punzado y/o usando un filtro de cola en el fondo de la bomba. * Usar algún tipo de ancla de gas. * Reemplazar la bomba por una de mayor desplazamiento. * Correr la bomba a velocidades más bajas para evitar desgastes prematuros y acortamiento de la vida útil de la bomba.
La productividad del pozo es inferior a la esperada	* Verificar el nivel de fluido. * Reducir la velocidad de la bomba. * Monitorear los cambios en la eficiencia volumétrica. * Comparar con las curvas de comportamiento de la bomba.
Existen altas pérdidas por fricción por el uso de centralizadores	* Replantear la necesidad de centralizadores. Si hay disponibles, usar otro tipo de centralizador. * Reemplazar la bomba por otra que pueda girar más lento sin centralizadores. * Cambiar la tubería si es posible.
El estator está gastado	* Sacar la bomba. Llevarla a un banco de ensayos y reemplazarla de ser necesario.
La admisión de la bomba está tapada	* Levantar el rotor fuera del estator, desplazar fluido por el tubing para limpiar el estator, re-espaciar, poner en producción y revisar la misma.

Anexo 3 Problemas típicos en BCP (Caso 2) [24]

Caso 2: Caudal intermitente. Baja eficiencia volumétrica. Velocidad normal. Consumo dentro del límite esperado.	
Causa probable	Acción recomendada
Bajo nivel de sumergencia	<ul style="list-style-type: none"> * Verificar el nivel. Bajar la velocidad de la bomba. * Asegurar que la velocidad no pase de 200 RPM. Si es necesario cambiar la bomba para cumplir los requisitos de producción.
Valores altos de RGP	<ul style="list-style-type: none"> * Proveer medios para las anclas de gas natural, instalando la bomba por debajo del punzado y/o usando un filtro de cola en el fondo de la bomba. Usar algún tipo de ancla de gas. * Reemplazar la bomba por una de mayor desplazamiento. Correr la bomba a velocidades más bajas para evitar desgastes prematuros y acortamiento de la vida de la bomba.
Bomba dañada o sub-diseñada	<ul style="list-style-type: none"> * Sacar la bomba. Revisarla en el banco para poder usarla en otra aplicación. * Verificar los requerimientos hidráulicos de la instalación. * Reemplazar la bomba por otra de mayor capacidad de presión y caudal para poder bajar las RPM.

Anexo 3 Problemas típicos en BCP (Caso 3) [24]

Caso 3: Caudal intermitente. Pobre eficiencia volumétrica. Velocidad más baja que la normal. Consumo más alto que el esperado.	
Causa probable	Acción recomendada
Mal espaciado. Rotor tocando en el niple de paro	* Sacar el rotor. Re-espaciar. Re-arrancar, revisar todos los parámetros.
Rotor aprisionado por exceso de temperatura o ataque químico	* Sacar la bomba. Revisar la temperatura de fondo. Seleccionar rotor más pequeño. * Verificar el análisis químico del fluido. * Si es necesario, cambiar la formulación del elastómero.
Rotor aprisionado por sólidos	* Levantar el rotor y lavar el estator.

Anexo 3 Problemas típicos en BCP (Caso 4) [24]

Caso 4: Sin producción. Pérdida de velocidad gradual. Consumo más alto que el esperado.	
Causa probable	Acción recomendada
Mal espaciado. El rotor está en contacto con el niple de paro	* Levantar el rotor. Re-espaciar. Re-arrancar. Revisar todos los parámetros. * Cambiar la bomba si es necesario.
Elastómero hinchado aumenta la fricción con el rotor	* Sacar la bomba. Verificar la temperatura de fondo. * Seleccionar un rotor más pequeño si es necesario. * Analizar el fluido. * Cambiar la composición del elastómero para cumplir con las condiciones de fondo.
Existe una alta interferencia entre el rotor y estator	* Reemplazar la bomba para otra capacidad de presión y caudal con distinto ajuste de compresión. * Seleccionar un rotor más pequeño. * Monitorear el consumo.

Anexo 3 Problemas típicos en BCP (Caso 5) [24]

Caso 5: Sin producción. Velocidad normal. Consumo bajo.	
Causa probable	Acción recomendada
Rotación contraria	* Verificar el giro. Verificar si no hay pesca. Arrancar de nuevo.
El rotor no está insertado en el estator	* Verificar las medidas de instalación. * Re-espaciar. Re-arrancar. * Monitorear el caudal.
El estator y el rotor están dañados	* Revisar la profundidad de la bomba y comparar con la longitud de barras. * Revisar la presión. Cambiar partes de ser necesario.
Rotor o barras de pesca	* Profundizar la instalación. Re-espaciar. Sacar y reparar. * Cambiar la bomba.
El tubing no tiene hermeticidad	* Verificar nivel y presión. Sacar la columna de producción. Reparar la pesca.
El tubing está desenroscado o cortado	* Verificar el espaciado. * Sacar la sarta de barras y el tubing. * Reparar.

Anexo 3 Problemas típicos en BCP (Caso 6) ^[24]

Caso 6: El motor eléctrico se detiene. La corriente es más alta de lo esperado.	
Causa probable	Acción recomendada
La potencia del motor es baja para la aplicación	<ul style="list-style-type: none"> * Verificar la potencia a partir de la hidráulica de la instalación. * Comparar con la corriente de la instalación. * Cambiar a otro motor más adecuado.
Existe una falla en la línea de alimentación	<ul style="list-style-type: none"> * Verificar las fases en la línea. * Re-arrancar el sistema.
El rotor está bloqueado dentro del estator debido a arena o incrustaciones	<ul style="list-style-type: none"> * <i>Flush-by</i>/circular el pozo para tratar de limpiarlo.
Hay hinchamiento del estator debido a ataque químico o temperatura	<ul style="list-style-type: none"> * Verificar la elección del elastómero.

Anexo 3 Problemas típicos en BCP (Caso 7) ^[24]

Caso 7: Las pérdidas a través del sistema de sello permanecen altas, a pesar de haber ajustado el sello.	
Causa probable	Acción recomendada
Las empacaduras están gastadas	<ul style="list-style-type: none"> * Verificar el estado de las empacaduras. Reemplazarlas si es necesario.
La camisa de sacrificio esta gastada	<ul style="list-style-type: none"> * Verificar la camisa y reemplazarla si está dañada. Cambiar también las empacaduras.

Anexo 3 Problemas típicos en BCP (Caso 8) [24]

Caso 8: Las correas se rompen frecuentemente. Velocidad correcta. Corriente dentro de lo esperado.	
Causa probable	Acción recomendada
Existe un mal alineamiento entre correas y poleas	* Verificar la alineación y corregir si es necesario.
Las poleas están gastadas y/o rotas	* Verificar y cambiar si es necesario.
Las correas no son las adecuadas para la aplicación	* Verificar si el perfil es el correcto para la polea. * Reemplazar por el adecuado juego de correas o poleas. * Solicitar soporte técnico desde algún representante de la compañía fabricante.

Anexo 3 Problemas típicos en BCP (Caso 9) [24]

Caso 9: El nivel de aceite baja en un período de tiempo corto.	
Causa probable	Acción recomendada
El sistema de sello está dañado, gastado o mal ajustado	* Verificar el sello. Reemplazarlo si es necesario. * Completar el nivel de aceite. Arrancar y verificar si hay pérdidas.
El tapón de drenaje está suelto	* Reajustar el tapón.

Anexo 3 Problemas típicos en BCP (Caso 10) [24]

Caso 10: Existen pérdidas a través del sistema de sellado del vástago.	
Causa probable	Acción recomendada
El sistema de sellado está dañado o mal armado	* Revisar los elementos del empaque. Reemplazarlos si es necesario.
El sistema de empaquetado está suelto	* Verificar el ajuste. Reajustar.
El vástago usado tiene la zona de empaque gastada o dañada	* Revisar el vástago en la zona de sello. Cambiarlo si no es posible cambiar su posición sin variar el espaciado.

Anexo 3 Problemas típicos en BCP (Caso 11) [24]

Caso 11: La temperatura del aceite del cabezal es alta.	
Causa probable	Acción recomendada
El cabezal está girando a mayor velocidad que la recomendada para ese modelo	* Verificar la velocidad. Cambiar la relación de poleas para alcanzar la velocidad deseada, de acuerdo al desplazamiento de la bomba. * Cambiar el tipo de cabezal por una elección más adecuada a la aplicación. * Solicitar soporte técnico a un representante de la empresa fabricante.
La especificación del aceite no es la recomendada	* Verificar el aceite. Reemplazar de ser necesario.
El nivel de aceite está más alto que el recomendado	* Verificar el nivel de aceite y corregir de ser necesario.

Anexo 4: Tabla de especificaciones de BCP

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
1-1200	NOV MONO	0,06	1-750	NOV MONO	0,06
1-1950	NOV MONO	0,06	1-1500	NOV MONO	0,06
2-1500	NOV MONO	0,1	2-1200	NOV MONO	0,1
3-1200	NOV MONO	0,15	3-600	NOV MONO	0,15
4-750	NOV MONO	0,25	3-1800	NOV MONO	0,15
4-1500	NOV MONO	0,25	4-1200	NOV MONO	0,25
4-2400	NOV MONO	0,25	4-1950	NOV MONO	0,25
6-700	NOV MONO	0,44	4-3000	NOV MONO	0,25
6-1400	NOV MONO	0,44	6-1050	NOV MONO	0,44
6-2100	NOV MONO	0,44	6-1750	NOV MONO	0,44
7-1200	NOV MONO	0,45	7-750	NOV MONO	0,45
7-1800	NOV MONO	0,45	7-1600	NOV MONO	0,45
7-3000	NOV MONO	0,45	7-2400	NOV MONO	0,45
8-2250	NOV MONO	0,56	8-1500	NOV MONO	0,56
9-1000	NOV MONO	0,6	9-600	NOV MONO	0,6
9-1800	NOV MONO	0,6	9-1400	NOV MONO	0,6
10-1200	NOV MONO	0,65	10-750	NOV MONO	0,65
10-1800	NOV MONO	0,65	10-1500	NOV MONO	0,65
10-3000	NOV MONO	0,65	10-2400	NOV MONO	0,65
11-1800	NOV MONO	0,67	10-3800	NOV MONO	0,65
15-1050	NOV MONO	0,9	15-700	NOV MONO	0,9
15-1800	NOV MONO	0,9	15-1400	NOV MONO	0,9
15-2800	NOV MONO	0,9	15-2100	NOV MONO	0,9
16-1200 HTL	NOV MONO	0,94	16-600 HTL	NOV MONO	0,94
17-750	NOV MONO	0,95	16-1800 HTL	NOV MONO	0,94
17-1500	NOV MONO	0,95	17-1200	NOV MONO	0,95
17-2400	NOV MONO	0,95	17-1800	NOV MONO	0,95
17-3600	NOV MONO	0,95	17-3000	NOV MONO	0,95
21-660	NOV MONO	1,31	21-440	NOV MONO	1,31
21-1100	NOV MONO	1,31	21-880	NOV MONO	1,31
22-660 ML	NOV MONO	1,35	21-1320	NOV MONO	1,31
22-2400 ML	NOV MONO	1,35	22-1500 ML	NOV MONO	1,35

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
23-750	NOV MONO	1,44	22-3000 ML	NOV MONO	1,35
23-1500	NOV MONO	1,44	23-1200	NOV MONO	1,44
23-2400	NOV MONO	1,44	23-1800	NOV MONO	1,44
23-3800	NOV MONO	1,44	23-3000	NOV MONO	1,44
24-900	NOV MONO	1,5	24-60	NOV MONO	1,5
24-1500	NOV MONO	1,5	24-1200	NOV MONO	1,5
24-2100	NOV MONO	1,5	24-1800	NOV MONO	1,5
27-1200	NOV MONO	1,7	27-800	NOV MONO	1,7
27-1800	NOV MONO	1,7	27-1600	NOV MONO	1,7
27-3000	NOV MONO	1,7	27-2400	NOV MONO	1,7
32-600	NOV MONO	2	27-3800	NOV MONO	1,7
32-1200	NOV MONO	2	32-900	NOV MONO	2
32-1800	NOV MONO	2	32-1500	NOV MONO	2
33-750	NOV MONO	2,05	32-2100	NOV MONO	2
33-1500	NOV MONO	2,05	33-1200	NOV MONO	2,05
33-2400	NOV MONO	2,05	33-1900	NOV MONO	2,05
39-1125	NOV MONO	2,45	39-750	NOV MONO	2,45
39-1875	NOV MONO	2,45	39-1500	NOV MONO	2,45
40-1200 ML	NOV MONO	2,5	40-750 ML	NOV MONO	2,5
44-1000	NOV MONO	2,75	40-1500 ML	NOV MONO	2,5
44-2000	NOV MONO	2,75	44-1500	NOV MONO	2,75
44-3000	NOV MONO	2,75	44-2400	NOV MONO	2,75
45-800 HTL	NOV MONO	2,8	45-400 HTL	NOV MONO	2,8
45-1600 HTL	NOV MONO	2,8	45-1200 HTL	NOV MONO	2,8
47-1125	NOV MONO	2,95	47-750	NOV MONO	2,95
47-1875	NOV MONO	2,95	47-1500	NOV MONO	2,95
54-600	NOV MONO	3,4	47-2250	NOV MONO	2,95
54-1200	NOV MONO	3,4	54-900	NOV MONO	3,4
54-1800	NOV MONO	3,4	54-1500	NOV MONO	3,4
64-900	NOV MONO	3,4	64-600	NOV MONO	3,4
64-1500	NOV MONO	3,4	64-1200	NOV MONO	3,4
70-900	NOV MONO	4,4	64-1800	NOV MONO	3,4
70-1800	NOV MONO	4,4	70-1350	NOV MONO	4,4
74-800	NOV MONO	4,95	74-200	NOV MONO	4,95
74-1200	NOV MONO	4,95	74-1000	NOV MONO	4,95
75-750	NOV MONO	5	75-500	NOV MONO	5

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
75-1250	NOV MONO	5	75-1000	NOV MONO	5
80-800	NOV MONO	5,05	75-1500	NOV MONO	5
80-1200	NOV MONO	5,05	80-1000	NOV MONO	5,05
80-1800	NOV MONO	5,05	80-1600	NOV MONO	5,05
96-800	NOV MONO	6	96-520	NOV MONO	6
96-1300	NOV MONO	6	96-1040	NOV MONO	6
98-790	NOV MONO	6,15	96-1560	NOV MONO	6
98-1580	NOV MONO	6,15	98-1200	NOV MONO	6,15
110-400	NOV MONO	6,85	98-1800	NOV MONO	6,15
110-1000	NOV MONO	6,85	110-800	NOV MONO	6,85
111-800	NOV MONO	6,9	110-1200	NOV MONO	6,85
111-1600	NOV MONO	6,9	111-1200	NOV MONO	6,9
120-400	NOV MONO	7,55	111-1800	NOV MONO	6,9
120-800	NOV MONO	7,55	120-600	NOV MONO	7,55
120-1200	NOV MONO	7,55	120-1000	NOV MONO	7,55
125-1200 ML	NOV MONO	7,85	125-700 ML	NOV MONO	7,85
130-650	NOV MONO	8,2	125-1500 ML	NOV MONO	7,85
130-1625	NOV MONO	8,2	130-1300	NOV MONO	8,2
150-360	NOV MONO	9,5	130-1950	NOV MONO	8,2
150-900	NOV MONO	9,5	150-720	NOV MONO	9,5
155-150 HTL	NOV MONO	9,75	150-1100	NOV MONO	9,5
155-600 HTL	NOV MONO	9,75	155-450 HTL	NOV MONO	9,75
155-900 HTL	NOV MONO	9,75	155-750 HTL	NOV MONO	9,75
165-1350	NOV MONO	10,3	165-900	NOV MONO	10,3
200-900	NOV MONO	12,6	165-1800	NOV MONO	10,3
200-1800	NOV MONO	12,6	200-1350	NOV MONO	12,6
206-900	NOV MONO	13	206-600	NOV MONO	13
206-1500	NOV MONO	13	206-1200	NOV MONO	13
334-1280	NOV MONO	21	334-640	NOV MONO	21
82-1200 MTM	NOV MONO	5,2	334-1500	NOV MONO	21
218-700 MTM	NOV MONO	13,7	112-1200 MTM	NOV MONO	7

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
4-900	NOV MONOFLO	0,25	4-600	NOV MONOFLO	0,25
4-1800	NOV MONOFLO	0,25	4-1200	NOV MONOFLO	0,25
10-900	NOV MONOFLO	0,6	10-600	NOV MONOFLO	0,6
10-1500	NOV MONOFLO	0,6	10-1200	NOV MONOFLO	0,6
10-2400	NOV MONOFLO	0,6	10-1800	NOV MONOFLO	0,6
15-700	NOV MONOFLO	0,95	13-1500	NOV MONOFLO	0,8
15-1800	NOV MONOFLO	0,95	15-1400	NOV MONOFLO	0,95
17-750	NOV MONOFLO	1,07	15-2000	NOV MONOFLO	0,95
24-600	NOV MONOFLO	1,5	17-1500	NOV MONOFLO	1,07
24-1200	NOV MONOFLO	1,5	24-900	NOV MONOFLO	1,5
24-1800	NOV MONOFLO	1,5	24-1500	NOV MONOFLO	1,5
32-900	NOV MONOFLO	2	32-600	NOV MONOFLO	2
32-1500	NOV MONOFLO	2	32-1200	NOV MONOFLO	2
40-750	NOV MONOFLO	2,5	32-1800	NOV MONOFLO	2
40-1500	NOV MONOFLO	2,5	40-1125	NOV MONOFLO	2,5
44-600	NOV MONOFLO	2,8	40-1875	NOV MONOFLO	2,5
44-1400	NOV MONOFLO	2,8	44-900	NOV MONOFLO	2,8
44-2100	NOV MONOFLO	2,8	44-1800	NOV MONOFLO	2,8
54-900	NOV MONOFLO	3,4	54-600	NOV MONOFLO	3,4
54-1500	NOV MONOFLO	3,4	54-1200	NOV MONOFLO	3,4
64-520	NOV MONOFLO	4	54-1800	NOV MONOFLO	3,4
64-1040	NOV MONOFLO	4	64-800	NOV MONOFLO	4

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
64-1560	NOV MONOFLO	4	64-1300	NOV MONOFLO	4
70-1200	NOV MONOFLO	4,4	70-900	NOV MONOFLO	4,4
70-1800	NOV MONOFLO	4,4	70-1500	NOV MONOFLO	4,4
80-1200	NOV MONOFLO	5	80-800	NOV MONOFLO	5
80-1800	NOV MONOFLO	5	80-1600	NOV MONOFLO	5
96-1040	NOV MONOFLO	6	96-800	NOV MONOFLO	6
120-400	NOV MONOFLO	7,55	96-1300	NOV MONOFLO	6
120-800	NOV MONOFLO	7,55	120-600	NOV MONOFLO	7,55
120-1200	NOV MONOFLO	7,55	120-1000	NOV MONOFLO	7,55
98-1580	NOV MONOFLO	6,15	98-1350	NOV MONOFLO	6,15
130-1300	NOV MONOFLO	8,2	110-1600	NOV MONOFLO	6,9
150-600	NOV MONOFLO	9,4	130-1625	NOV MONOFLO	8,2
150-900	NOV MONOFLO	9,4	150-750	NOV MONOFLO	9,4
165-1800	NOV MONOFLO	10,3	165-1200	NOV MONOFLO	10,3
200-1800	NOV MONOFLO	12,6	200-1200	NOV MONOFLO	12,6
25-3100	NOV MONOFLO	0,25	25-2000	NOV MONOFLO	0,25
25-6000	NOV MONOFLO	0,25	25-4100	NOV MONOFLO	0,25
60-3100	NOV MONOFLO	0,6	60-2000	NOV MONOFLO	0,6
60-5000	NOV MONOFLO	0,6	60-4100	NOV MONOFLO	0,6
60-8000	NOV MONOFLO	0,6	60-6000	NOV MONOFLO	0,6
95-2400	NOV MONOFLO	0,95	80-5000	NOV MONOFLO	0,8
95-6000	NOV MONOFLO	0,95	95-4600	NOV MONOFLO	0,95

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
107-2500	NOV MONOFLO	1,07	95-6600	NOV MONOFLO	0,95
150-2000	NOV MONOFLO	1,5	107-5000	NOV MONOFLO	1,07
150-4100	NOV MONOFLO	1,5	150-3100	NOV MONOFLO	1,5
150-6000	NOV MONOFLO	1,5	150-5000	NOV MONOFLO	1,5
200-3100	NOV MONOFLO	2	200-2000	NOV MONOFLO	2
200-5000	NOV MONOFLO	2	200-4100	NOV MONOFLO	2
250-2700	NOV MONOFLO	2,5	200-6000	NOV MONOFLO	2
250-5000	NOV MONOFLO	2,5	250-3750	NOV MONOFLO	2,5
280-2000	NOV MONOFLO	2,8	250-6250	NOV MONOFLO	2,5
280-4600	NOV MONOFLO	2,8	280-3100	NOV MONOFLO	2,8
280-6900	NOV MONOFLO	2,8	280-6000	NOV MONOFLO	2,8
340-3100	NOV MONOFLO	3,4	340-2000	NOV MONOFLO	3,4
340-5000	NOV MONOFLO	3,4	340-4100	NOV MONOFLO	3,4
400-1700	NOV MONOFLO	4	340-6000	NOV MONOFLO	3,4
400-3500	NOV MONOFLO	4	400-2700	NOV MONOFLO	4
400-5350	NOV MONOFLO	4	400-4400	NOV MONOFLO	4
440-4100	NOV MONOFLO	4,4	440-3100	NOV MONOFLO	4,4
440-6000	NOV MONOFLO	4,4	440-5000	NOV MONOFLO	4,4
500-4100	NOV MONOFLO	5	500-2700	NOV MONOFLO	5
500-6000	NOV MONOFLO	5	500-5300	NOV MONOFLO	5
600-3500	NOV MONOFLO	6	600-2700	NOV MONOFLO	6
755-1300	NOV MONOFLO	7,55	600-4400	NOV MONOFLO	6

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
755-2700	NOV MONOFLO	7,55	755-2100	NOV MONOFLO	7,55
755-4100	NOV MONOFLO	7,55	755-3400	NOV MONOFLO	7,55
615-5200	NOV MONOFLO	6,15	615-4430	NOV MONOFLO	6,15
820-4400	NOV MONOFLO	8,2	690-5300	NOV MONOFLO	6,9
940-2000	NOV MONOFLO	9,4	820-5380	NOV MONOFLO	8,2
940-3100	NOV MONOFLO	9,4	940-2500	NOV MONOFLO	9,4
1030-6000	NOV MONOFLO	10,3	1030-4100	NOV MONOFLO	10,3
1260-6000	NOV MONOFLO	12,6	1260-4100	NOV MONOFLO	12,6
40.40-1200	WEATHERFOR D	2,6	24.40-1200	WEATHERFOR D	2,4
20.40-1200	WEATHERFOR D	4,34	32.65-3000	WEATHERFOR D	6
130-1600	WEATHERFOR D	8,15	24.40-1500	WEATHERFOR D	3
161-1600	WEATHERFOR D	10,11	161-1800	WEATHERFOR D	10,12
28.40-2100	WEATHERFOR D	4,2	20.40-2100	WEATHERFOR D	4,2
22.40-2500	WEATHERFOR D	5	28.45-700	WEATHERFOR D	1,4
98-1600	WEATHERFOR D	6,15	21.65-3000	WEATHERFOR D	6,7
50-H-400	R&M MOYNO	4	60-N-340	R&M MOYNO	3,27
50-N-275	R&M MOYNO	2,75	50-N-170	R&M MOYNO	1,7
62-P-620	R&M MOYNO	6,2	80-N-275	R&M MOYNO	2,768
50-P-1000	R&M MOYNO	12	50-H-800	R&M MOYNO	8
50-N-340	R&M MOYNO	3,4	50-P-1300	R&M MOYNO	12,9
0200-D- 2300	LIFTEQ	2	50-N-195	R&M MOYNO	1,95
270-G-2300	LIFTEQ	2,7	270-D-2300	LIFTEQ	2,7
500-G-1800	LIFTEQ	5	350-D-2300	LIFTEQ	3,5
700-G-2300	LIFTEQ	7	500-G-2300	LIFTEQ	5
60-TP-1300	PCM	0,83	900-G-2300	LIFTEQ	9

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
750-TP-1200	PCM	9,43	80-TP-1600	PCM	1
900-TP-1500	PCM	11,32	400-TP-1350	PCM	5
430-TP-2000	PCM	5,42	120-TP-2000	PCM	1,5
300-TP-1800	PCM	3,8	200-TP-1800	PCM	2,5
NTZ400*15 0ST62	NETZSCH	4	580-TP-1600	PCM	7,3
NTZ550*15 0ST176	NETZSCH	12	NTZ550*15 0ST145	NETZSCH	9
22-1800	TIERRA ALTA	1,38	NTZ550*15 0ST128	NETZSCH	8,05
55-1600	TIERRA ALTA	3,46	45-1800	TIERRA ALTA	2,83
105-1600	TIERRA ALTA	6,6	85-1600	TIERRA ALTA	3,5
80-1600	NATIONAL	5	170-1600	TIERRA ALTA	10
110-1600	NATIONAL	6,9	98-1580	NATIONAL	6
3E2400	PCM MOINEAU	0,19	3E1200	PCM MOINEAU	0,19
6E1300	PCM MOINEAU	0,3225	6E600	PCM MOINEAU	0,3225
16E1200	PCM MOINEAU	1,015	6E2000	PCM MOINEAU	0,3225
13E1300	PCM MOINEAU	0,77	16E1600	PCM MOINEAU	1,015
13E2600	PCM MOINEAU	0,77	13E2000	PCM MOINEAU	0,77
22E1800	PCM MOINEAU	1,2925	22E1200	PCM MOINEAU	1,2925
24E2000	PCM MOINEAU	1,4525	48E900	PCM MOINEAU	2,85
32E1500	PCM MOINEAU	1,9	24E2600	PCM MOINEAU	1,4525
40E600	PCM MOINEAU	2,3375	32E2200	PCM MOINEAU	1,9
40E1800	PCM MOINEAU	2,3375	40E1200	PCM MOINEAU	2,3375
63E1200	PCM MOINEAU	3,5825	63E800	PCM MOINEAU	3,5825
45E2400	PCM MOINEAU	2,6875	45E1600	PCM MOINEAU	2,6875
60E1800	PCM MOINEAU	3,61	60E1200	PCM MOINEAU	3,61
80E900	PCM MOINEAU	4,8175	60E2400	PCM MOINEAU	3,61
80E1800	PCM MOINEAU	4,8175	80E1350	PCM MOINEAU	4,8175
120E900	PCM MOINEAU	7,1625	120E600	PCM MOINEAU	7,1625
160E600	PCM MOINEAU	9,5475	120E1200	PCM MOINEAU	7,1625
116E1600	PCM MOINEAU	6,8875	86E2000	PCM MOINEAU	5,1575
200E860	PCM MOINEAU	12,54	150E1200	PCM MOINEAU	8,9975
80V1000	PCM MOINEAU	3,75	185E1500	PCM MOINEAU	11,02
220V750	PCM MOINEAU	10	110V750	PCM MOINEAU	5,25

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
15TP1200	PCM MOINEAU	0,19	220V1000	PCM MOINEAU	10
30TP600	PCM MOINEAU	0,3225	15TP2400	PCM MOINEAU	0,19
30TP2000	PCM MOINEAU	0,3225	30TP1300	PCM MOINEAU	0,3225
80TP1600	PCM MOINEAU	1,015	80TP1200	PCM MOINEAU	1,015
60TP2000	PCM MOINEAU	0,77	60TP1300	PCM MOINEAU	0,77
100TP1200	PCM MOINEAU	1,2925	60TP2600	PCM MOINEAU	0,77
240TP900	PCM MOINEAU	2,85	100TP1800	PCM MOINEAU	1,2925
120TP2600	PCM MOINEAU	1,4525	120TP2000	PCM MOINEAU	1,4525
160TP2200	PCM MOINEAU	1,9	160TP1500	PCM MOINEAU	1,9
200TP1200	PCM MOINEAU	2,3375	20TP600	PCM MOINEAU	2,3375
300TP800	PCM MOINEAU	3,5825	200TP1800	PCM MOINEAU	2,3375
225TP1600	PCM MOINEAU	2,6875	300TP1210	PCM MOINEAU	3,5825
300TP1200	PCM MOINEAU	3,61	225TP2400	PCM MOINEAU	2,6875
30TP2400	PCM MOINEAU	3,61	300TP1800	PCM MOINEAU	3,61
400TP1350	PCM MOINEAU	4,8175	400TP900	PCM MOINEAU	4,8175
600TP600	PCM MOINEAU	7,1625	400TP1800	PCM MOINEAU	4,8175
600TP1200	PCM MOINEAU	7,1625	600TP900	PCM MOINEAU	7,1625
430TP2000	PCM MOINEAU	5,1575	800TP600	PCM MOINEAU	9,5475
750TP1200	PCM MOINEAU	8,9975	580TP1600	PCM MOINEAU	6,8875
900TP1500	PCM MOINEAU	11,02	1000TP860	PCM MOINEAU	12,54
550MET750	PCM MOINEAU	5,25	400MET100 0	PCM MOINEAU	3,75
110MET100 0	PCM MOINEAU	10	1100MET50 0	PCM MOINEAU	10
9-N-1	R&M MOYNO	0,072	6-N-1	R&M MOYNO	0,072
18-N-1	R&M MOYNO	0,072	12-N-1	R&M MOYNO	0,072
9-N-2.3	R&M MOYNO	0,168	6-N-2.3	R&M MOYNO	0,168
18-N-2	R&M MOYNO	0,168	12-N-2.3	R&M MOYNO	0,168
9-S-4	R&M MOYNO	0,268	6-N-4	R&M MOYNO	0,268
15-S-4	R&M MOYNO	0,268	12-S-4	R&M MOYNO	0,268
24-S-4	R&M MOYNO	0,268	18-S-4	R&M MOYNO	0,268
9-S-4.5	R&M MOYNO	0,295	6-S-4.5	R&M MOYNO	0,295
15-S-4.5	R&M MOYNO	0,295	12-S-4.5	R&M MOYNO	0,295
6-S-7	R&M MOYNO	0,495	18-S-4.5	R&M MOYNO	0,295
12-S-7	R&M MOYNO	0,495	9-S-7	R&M MOYNO	0,495
18-S-7	R&M MOYNO	0,495	15-S-7	R&M MOYNO	0,495
5-S-7.3	R&M MOYNO	0,47	24-S-7	R&M MOYNO	0,495

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
9-S-7.3	R&M MOYNO	0,47	6-S-7.3	R&M MOYNO	0,47
14-S-7.3	R&M MOYNO	0,47	11-S-7.3	R&M MOYNO	0,47
6-H-10	R&M MOYNO	0,67	13-S-9	R&M MOYNO	0,587
12-H-10	R&M MOYNO	0,67	9-H-10	R&M MOYNO	0,67
18-H-10	R&M MOYNO	0,67	15-H-10	R&M MOYNO	0,67
30-H-10	R&M MOYNO	0,67	24-H-10	R&M MOYNO	0,67
6-ML-17	R&M MOYNO	0,988	36-H-10	R&M MOYNO	0,67
3-N-17	R&M MOYNO	0,988	12-ML-17	R&M MOYNO	0,988
9-N-17	R&M MOYNO	0,988	6-N-17	R&M MOYNO	0,988
15-N-17	R&M MOYNO	0,988	12-N-17	R&M MOYNO	0,988
24-N-17	R&M MOYNO	0,988	18-N-17	R&M MOYNO	0,988
36-N-17	R&M MOYNO	0,988	30-N-17	R&M MOYNO	0,988
16-N-15	R&M MOYNO	1,027	14-N-15	R&M MOYNO	1,027
24-N-15	R&M MOYNO	1,027	18-N-15	R&M MOYNO	1,027
3-T-22	R&M MOYNO	1,452	28-N-15	R&M MOYNO	1,027
12-T-22	R&M MOYNO	1,452	6-T-22	R&M MOYNO	1,452
24-T-22	R&M MOYNO	1,452	18-T-22	R&M MOYNO	1,452
15-S-23	R&M MOYNO	1,44	12-S-23	R&M MOYNO	1,44
24-S-23	R&M MOYNO	1,44	18-S-23	R&M MOYNO	1,44
36-S-23	R&M MOYNO	1,44	30-S-23	R&M MOYNO	1,44
15-S-27	R&M MOYNO	1,7	12-S-27	R&M MOYNO	1,7
24-S-27	R&M MOYNO	1,7	18-S-27	R&M MOYNO	1,7
36-S-27	R&M MOYNO	1,7	30-S-27	R&M MOYNO	1,7
12-N-31	R&M MOYNO	1,982	9-N-31	R&M MOYNO	1,982
18-N-31	R&M MOYNO	1,982	15-N-31	R&M MOYNO	1,982
3-H-32	R&M MOYNO	2,05	24-N-31	R&M MOYNO	1,982
9-H-32	R&M MOYNO	2,05	6-H-32	R&M MOYNO	2,05
15-H-32	R&M MOYNO	2,05	12-H-32	R&M MOYNO	2,05
7-T-40	R&M MOYNO	2,485	18-H-32	R&M MOYNO	2,05
14-T-40	R&M MOYNO	2,485	10-T-40	R&M MOYNO	2,485
12-ML-44	R&M MOYNO	2,801	8-ML-44	R&M MOYNO	2,801
8-N-44	R&M MOYNO	2,801	16-ML-44	R&M MOYNO	2,801
16-N-44	R&M MOYNO	2,801	12-N-44	R&M MOYNO	2,801
24-N-44	R&M MOYNO	2,801	20-N-44	R&M MOYNO	2,801
8-S-44	R&M MOYNO	2,801	27-N-44	R&M MOYNO	2,801
16-S-44	R&M MOYNO	2,801	12-S-44	R&M MOYNO	2,801

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
24-S-44	R&M MOYNO	2,801	20-S-44	R&M MOYNO	2,801
10-N-49	R&M MOYNO	3,109	27-S-44	R&M MOYNO	2,801
30-N-49	R&M MOYNO	3,109	20-N-49	R&M MOYNO	3,109
16-T-50	R&M MOYNO	3,204	8-T-50	R&M MOYNO	3,204
6-N-52	R&M MOYNO	3,48	3-N-52	R&M MOYNO	3,48
12-N-52	R&M MOYNO	3,48	9-N-52	R&M MOYNO	3,48
18-N-52	R&M MOYNO	3,48	15-N-52	R&M MOYNO	3,48
6-S-52	R&M MOYNO	3,48	3-S-52	R&M MOYNO	3,48
12-S-52	R&M MOYNO	3,48	9-S-52	R&M MOYNO	3,48
18-S-52	R&M MOYNO	3,48	15-S-52	R&M MOYNO	3,48
7-H-64	R&M MOYNO	4,086	5-H-64	R&M MOYNO	4,086
12-H-64	R&M MOYNO	4,086	10-H-64	R&M MOYNO	4,086
4-H-75	R&M MOYNO	5,116	15-H-64	R&M MOYNO	4,086
8-H-75	R&M MOYNO	5,116	6-H-75	R&M MOYNO	5,116
12-H-75	R&M MOYNO	5,116	10-H-75	R&M MOYNO	5,116
4-ML-75	R&M MOYNO	5,116	2-ML-75	R&M MOYNO	5,116
4-S-75	R&M MOYNO	5,116	12-ML-75	R&M MOYNO	5,116
8-S-75	R&M MOYNO	5,116	6-S-75	R&M MOYNO	5,116
12-S-75	R&M MOYNO	5,116	10-S-75	R&M MOYNO	5,116
8-H-93	R&M MOYNO	5,894	4-H-93	R&M MOYNO	5,894
12-H-93	R&M MOYNO	5,894	10-H-93	R&M MOYNO	5,894
10-S-93	R&M MOYNO	5,894	8-S-93	R&M MOYNO	5,894
6-P-98	R&M MOYNO	6,549	12-S-93	R&M MOYNO	5,894
15-P-98	R&M MOYNO	6,549	12-P-98	R&M MOYNO	6,549
3-H-110	R&M MOYNO	7,071	18-P-98	R&M MOYNO	6,549
6-H-110	R&M MOYNO	7,071	5-H-110	R&M MOYNO	7,071
9-H-110	R&M MOYNO	7,071	8-H-110	R&M MOYNO	7,071
5-S-110	R&M MOYNO	7,071	3-S-110	R&M MOYNO	7,071
8-S-110	R&M MOYNO	7,071	6-S-110	R&M MOYNO	7,071
6-TS-125	R&M MOYNO	7,985	9-S-110	R&M MOYNO	7,071
12-TS-125	R&M MOYNO	7,985	9-TS-125	R&M MOYNO	7,985
9-H-125	R&M MOYNO	8,071	6-H-125	R&M MOYNO	8,071
15-H-125	R&M MOYNO	8,071	12-H-125	R&M MOYNO	8,071
5-H-150	R&M MOYNO	9,852	3-H-150	R&M MOYNO	9,852
7-H-150	R&M MOYNO	9,852	6-H-150	R&M MOYNO	9,852
9-H-150	R&M MOYNO	9,852	8-H-150	R&M MOYNO	9,852

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
11-P-159	R&M MOYNO	10,261	7-P-159	R&M MOYNO	10,261
18-P-159	R&M MOYNO	10,261	14-P-159	R&M MOYNO	10,261
7-M-206	R&M MOYNO	13,719	22-P-159	R&M MOYNO	10,261
9-P-206	R&M MOYNO	13,867	6-P-206	R&M MOYNO	13,867
15-P-206	R&M MOYNO	13,867	12-P-206	R&M MOYNO	13,867
15-P-334	R&M MOYNO	21,542	6-P-334	R&M MOYNO	21,542
30-N-006	R&M MOYNO	0,072	20-N-006	R&M MOYNO	0,072
60-N-006	R&M MOYNO	0,072	40-N-006	R&M MOYNO	0,072
30-N-015	R&M MOYNO	0,168	20-N-015	R&M MOYNO	0,168
60-N-015	R&M MOYNO	0,168	40-N-015	R&M MOYNO	0,168
30-S-025	R&M MOYNO	0,268	20-S-025	R&M MOYNO	0,268
50-S-025	R&M MOYNO	0,268	40-S-025	R&M MOYNO	0,268
80-S-025	R&M MOYNO	0,268	60-S-025	R&M MOYNO	0,268
30-S-28	R&M MOYNO	0,295	20-S-28	R&M MOYNO	0,295
50-S-28	R&M MOYNO	0,295	40-S-28	R&M MOYNO	0,295
20-S-045	R&M MOYNO	0,495	60-S-28	R&M MOYNO	0,295
40-S-045	R&M MOYNO	0,495	30-S-045	R&M MOYNO	0,495
60-S-045	R&M MOYNO	0,495	50-S-045	R&M MOYNO	0,495
15-S-046	R&M MOYNO	0,47	80-S-045	R&M MOYNO	0,495
30-S-046	R&M MOYNO	0,47	23-S-046	R&M MOYNO	0,47
45-S-046	R&M MOYNO	0,47	37-S-046	R&M MOYNO	0,47
20-H-065	R&M MOYNO	0,67	42-S-060	R&M MOYNO	0,587
40-H-065	R&M MOYNO	0,67	30-H-065	R&M MOYNO	0,67
60-H-065	R&M MOYNO	0,67	50-H-065	R&M MOYNO	0,67
100-H-065	R&M MOYNO	0,67	80-H-065	R&M MOYNO	0,67
20-ML-095	R&M MOYNO	0,988	120-H-065	R&M MOYNO	0,67
10-N-095	R&M MOYNO	0,988	40-ML-095	R&M MOYNO	0,988
30-N-095	R&M MOYNO	0,988	20-N-095	R&M MOYNO	0,988
50-N-095	R&M MOYNO	0,988	40-N-095	R&M MOYNO	0,988
80-N-095	R&M MOYNO	0,988	60-N-095	R&M MOYNO	0,988
120-N-095	R&M MOYNO	0,988	100-N-095	R&M MOYNO	0,988
54-N-100	R&M MOYNO	1,027	47-N-100	R&M MOYNO	1,027
80-N-100	R&M MOYNO	1,027	60-N-100	R&M MOYNO	1,027
10-T-135	R&M MOYNO	1,452	94-N-100	R&M MOYNO	1,027
40-T-135	R&M MOYNO	1,452	20-T-135	R&M MOYNO	1,452
80-T-135	R&M MOYNO	1,452	60-T-135	R&M MOYNO	1,452

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
50-S-144	R&M MOYNO	1,44	40-S-144	R&M MOYNO	1,44
80-S-144	R&M MOYNO	1,44	60-S-144	R&M MOYNO	1,44
120-S-144	R&M MOYNO	1,44	100-S-144	R&M MOYNO	1,44
50-S-170	R&M MOYNO	1,7	40-S-170	R&M MOYNO	1,7
80-S-170	R&M MOYNO	1,7	60-S-170	R&M MOYNO	1,7
120-S-170	R&M MOYNO	1,7	100-S-170	R&M MOYNO	1,7
40-N-195	R&M MOYNO	1,982	30-N-195	R&M MOYNO	1,982
60-N-195	R&M MOYNO	1,982	50-N-195	R&M MOYNO	1,982
9-H-200	R&M MOYNO	2,05	80-N-195	R&M MOYNO	1,982
30-H-200	R&M MOYNO	2,05	20-H-200	R&M MOYNO	2,05
50-H-200	R&M MOYNO	2,05	40-H-200	R&M MOYNO	2,05
22-T-250	R&M MOYNO	2,485	60-H-200	R&M MOYNO	2,05
44-T-250	R&M MOYNO	2,485	33-T-250	R&M MOYNO	2,485
40-ML-275	R&M MOYNO	2,801	27-ML-275	R&M MOYNO	2,801
27-N-275	R&M MOYNO	2,801	54-ML-275	R&M MOYNO	2,801
54-N-275	R&M MOYNO	2,801	40-N-275	R&M MOYNO	2,801
80-N-275	R&M MOYNO	2,801	67-N-275	R&M MOYNO	2,801
27-S-275	R&M MOYNO	2,801	89-N-275	R&M MOYNO	2,801
54-S-275	R&M MOYNO	2,801	40-S-275	R&M MOYNO	2,801
80-S-275	R&M MOYNO	2,801	67-S-275	R&M MOYNO	2,801
33-N-310	R&M MOYNO	3,109	89-S-275	R&M MOYNO	2,801
100-N-310	R&M MOYNO	3,109	66-N-310	R&M MOYNO	3,109
54-T-315	R&M MOYNO	3,204	27-T-315	R&M MOYNO	3,204
20-N-340	R&M MOYNO	3,48	10-N-340	R&M MOYNO	3,48
40-N-340	R&M MOYNO	3,48	30-N-340	R&M MOYNO	3,48
60-N-340	R&M MOYNO	3,48	50-N-340	R&M MOYNO	3,48
20-S-340	R&M MOYNO	3,48	10-S-340	R&M MOYNO	3,48
40-S-340	R&M MOYNO	3,48	30-S-340	R&M MOYNO	3,48
60-S-340	R&M MOYNO	3,48	50-S-340	R&M MOYNO	3,48
22-H-400	R&M MOYNO	4,086	15-H-400	R&M MOYNO	4,086
40-H-400	R&M MOYNO	4,086	30-H-400	R&M MOYNO	4,086
13-H-500	R&M MOYNO	5,116	50-H-400	R&M MOYNO	4,086
27-H-500	R&M MOYNO	5,116	20-H-500	R&M MOYNO	5,116
40-H-500	R&M MOYNO	5,116	33-H-500	R&M MOYNO	5,116
13-ML-500	R&M MOYNO	5,116	7-ML-500	R&M MOYNO	5,116
13-S-500	R&M MOYNO	5,116	40-ML-500	R&M MOYNO	5,116

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
27-S-500	R&M MOYNO	5,116	20-S-500	R&M MOYNO	5,116
40-S-500	R&M MOYNO	5,116	33-S-500	R&M MOYNO	5,116
26-H-580	R&M MOYNO	5,894	13-H-580	R&M MOYNO	5,894
40-H-580	R&M MOYNO	5,894	33-H-580	R&M MOYNO	5,894
33-S-580	R&M MOYNO	5,894	26-S-580	R&M MOYNO	5,894
20-P-620	R&M MOYNO	6,549	40-S-580	R&M MOYNO	5,894
50-P-620	R&M MOYNO	6,549	40-P-620	R&M MOYNO	6,549
10-H-685	R&M MOYNO	7,071	60-P-620	R&M MOYNO	6,549
20-H-5685	R&M MOYNO	7,071	15-H-685	R&M MOYNO	7,071
30-H-685	R&M MOYNO	7,071	26-H-685	R&M MOYNO	7,071
15-S-685	R&M MOYNO	7,071	10-S-685	R&M MOYNO	7,071
26-S-685	R&M MOYNO	7,071	20-S-685	R&M MOYNO	7,071
20-TS-785	R&M MOYNO	7,985	30-S-685	R&M MOYNO	7,071
40-TS-785	R&M MOYNO	7,985	30-TS-785	R&M MOYNO	7,985
30-H-800	R&M MOYNO	8,071	20-H-800	R&M MOYNO	8,071
50-H-800	R&M MOYNO	8,071	40-H-800	R&M MOYNO	8,071
15-H-950	R&M MOYNO	9,852	10-H-950	R&M MOYNO	9,852
24-H-950	R&M MOYNO	9,852	20-H-950	R&M MOYNO	9,852
30-H-950	R&M MOYNO	9,852	27-H-950	R&M MOYNO	9,852
36-P-1000	R&M MOYNO	10,261	24-P-1000	R&M MOYNO	10,261
60-P-1000	R&M MOYNO	10,261	48-P-1000	R&M MOYNO	10,261
22-M-1300	R&M MOYNO	13,719	72-P-1000	R&M MOYNO	10,261
30-P-1300	R&M MOYNO	13,867	20-P-1300	R&M MOYNO	13,867
50-P-1300	R&M MOYNO	13,867	40-P-1300	R&M MOYNO	13,867
50-P-2100	R&M MOYNO	21,542	21-P-2100	R&M MOYNO	21,542
1-100	Weatherford	0,07	1-600	Weatherford	0,07
2-2000	Weatherford	0,12	1-1200	Weatherford	0,07
3-1000	Weatherford	0,2	3-600	Weatherford	0,2
4-2000	Weatherford	0,25	3-1200	Weatherford	0,16
5-1000	Weatherford	0,32	5-600	Weatherford	0,32
5-1200	Weatherford	0,32	5-1200	Weatherford	0,32
7-1000	Weatherford	0,44	7-600	Weatherford	0,44
11-1000	Weatherford	0,69	11-600	Weatherford	0,69
11-2000	Weatherford	0,69	11-1200	Weatherford	0,69
17-2000	Weatherford	1,07	14-1200	Weatherford	0,88
18-1000	Weatherford	1,13	18-600	Weatherford	1,13

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
24-2000	Weatherford	1,51	22-1200	Weatherford	1,38
28-1000	Weatherford	1,76	28-600	Weatherford	1,76
41-1000	Weatherford	2,58	41-600	Weatherford	2,58
41-2200	Weatherford	2,58	41-1700	Weatherford	2,58
54-600	Weatherford	3,4	41-2800	Weatherford	2,58
69-1200	Weatherford	4,34	55-1200	Weatherford	3,46
82-800	Weatherford	5,16	69-1400	Weatherford	4,34
1-1000	Weatherford	0,07	82-1500	Weatherford	5,16
7-600	Weatherford	0,44	1-2000	Weatherford	0,07
18-600	Weatherford	1,13	3-1200	Weatherford	0,2
9GP-20	Weatherford	0,07	7-1000	Weatherford	0,44
18GP-20	Weatherford	0,07	18-1000	Weatherford	1,13
9.20-100	Weatherford	0,2	14GP-20	Weatherford	0,07
18.20-80	Weatherford	0,16	28.20-60	Weatherford	0,12
9GP-30	Weatherford	0,32	14.20-100	Weatherford	0,2
18GP-30	Weatherford	0,32	28.25-125	Weatherford	0,25
9.25-200	Weatherford	0,44	14GP-30	Weatherford	0,32
9GP-40	Weatherford	0,69	18.25-150	Weatherford	0,32
18GP-40	Weatherford	0,69	14.25-200	Weatherford	0,44
18.35-400	Weatherford	0,88	14GP-40	Weatherford	0,69
9.35-500	Weatherford	1,13	28.35-300	Weatherford	0,69
18.40-600	Weatherford	1,38	28.40-500	Weatherford	1,07
9.40-800	Weatherford	1,76	14.35-500	Weatherford	1,13
9.45-1200	Weatherford	2,58	29.45-700	Weatherford	1,51
24.40-1200	Weatherford	2,58	14.40-800	Weatherford	1,76
40.40-1200	Weatherford	2,58	14.45-1200	Weatherford	2,58
18.40-1500	Weatherford	3,46	32.40-1200	Weatherford	2,58
20.40-2100	Weatherford	4,34	9.50-1700	Weatherford	3,4
22.40-2500	Weatherford	5,16	14.40-2100	Weatherford	4,34
28.20-35IM	Weatherford	0,07	11.40-2500	Weatherford	5,16
9.20-100IM	Weatherford	0,2	14.20-35IM	Weatherford	0,07
18.20-100IM	Weatherford	0,2	28.20-60IM	Weatherford	0,12
18.25-150IM	Weatherford	0,32	14.20-100IM	Weatherford	0,2
14.25-200IM	Weatherford	0,44	28.25-125IM	Weatherford	0,25

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
18.35-4001IM	Weatherford	0,88	9.25-200IM	Weatherford	0,44
14.35-500IM	Weatherford	1,13	28.35-300IM	Weatherford	0,69
25-B-1300	BAKER HUGHES CENTRILIFT	0,25	9.35-500IM	Weatherford	1,13
25-B-2300	BAKER HUGHES CENTRILIFT	0,25	25-B-0900	BAKER HUGHES CENTRILIFT	0,25
44-B-0900	BAKER HUGHES CENTRILIFT	0,44	25-B-1800	BAKER HUGHES CENTRILIFT	0,25
44-B-1800	BAKER HUGHES CENTRILIFT	0,44	25-B-2600	BAKER HUGHES CENTRILIFT	0,25
44-B-2600	BAKER HUGHES CENTRILIFT	0,44	44-B-1300	BAKER HUGHES CENTRILIFT	0,44
65-B-1300	BAKER HUGHES CENTRILIFT	0,65	44-B-2300	BAKER HUGHES CENTRILIFT	0,44
65-B-2300	BAKER HUGHES CENTRILIFT	0,65	65-B-0900	BAKER HUGHES CENTRILIFT	0,65
95-B-0900	BAKER HUGHES CENTRILIFT	0,95	65-B-1800	BAKER HUGHES CENTRILIFT	0,65
95-B-1800	BAKER HUGHES CENTRILIFT	0,95	65-B-2600	BAKER HUGHES CENTRILIFT	0,65
95-B-2600	BAKER HUGHES CENTRILIFT	0,95	95-B-1300	BAKER HUGHES CENTRILIFT	0,95
250-B-1300	BAKER HUGHES CENTRILIFT	2,5	95-B-2300	BAKER HUGHES CENTRILIFT	0,95
250-B-2300	BAKER HUGHES CENTRILIFT	2,5	250-B-0900	BAKER HUGHES CENTRILIFT	2,5
350-B-0900	BAKER HUGHES CENTRILIFT	3,5	250-B-1800	BAKER HUGHES CENTRILIFT	2,5

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
350-B-1800	BAKER HUGHES CENTRILIFT	3,5	250-B-2600	BAKER HUGHES CENTRILIFT	2,5
350-B-2600	BAKER HUGHES CENTRILIFT	3,5	350-B-1300	BAKER HUGHES CENTRILIFT	3,5
130-D-0900	BAKER HUGHES CENTRILIFT	1,3	350-B-2300	BAKER HUGHES CENTRILIFT	3,5
130-D-1800	BAKER HUGHES CENTRILIFT	1,3	110-D-2600	BAKER HUGHES CENTRILIFT	1,1
130-D-2600	BAKER HUGHES CENTRILIFT	1,3	130-D-1300	BAKER HUGHES CENTRILIFT	1,3
200-D-1300	BAKER HUGHES CENTRILIFT	2	130-D-2300	BAKER HUGHES CENTRILIFT	1,3
200-D-2300	BAKER HUGHES CENTRILIFT	2	200-D-0900	BAKER HUGHES CENTRILIFT	2
270-D-0900	BAKER HUGHES CENTRILIFT	2,7	200-D-1800	BAKER HUGHES CENTRILIFT	2
270-D-1800	BAKER HUGHES CENTRILIFT	2,7	200-D-2600	BAKER HUGHES CENTRILIFT	2
270-D-2600	BAKER HUGHES CENTRILIFT	2,7	270-D-1300	BAKER HUGHES CENTRILIFT	2,7
350-D-1300	BAKER HUGHES CENTRILIFT	3,5	270-D-2300	BAKER HUGHES CENTRILIFT	2,7
350-D-2300	BAKER HUGHES CENTRILIFT	3,5	350-D-0900	BAKER HUGHES CENTRILIFT	3,5
450-D-0900	BAKER HUGHES CENTRILIFT	4,5	350-D-1800	BAKER HUGHES CENTRILIFT	3,5
450-D-1800	BAKER HUGHES CENTRILIFT	4,5	350-D-2600	BAKER HUGHES CENTRILIFT	3,5

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
450-D-2600	BAKER HUGHES CENTRILIFT	4,5	450-D-1300	BAKER HUGHES CENTRILIFT	4,5
550-D-1300	BAKER HUGHES CENTRILIFT	5,5	450-D-2300	BAKER HUGHES CENTRILIFT	4,5
550-D-2300	BAKER HUGHES CENTRILIFT	5,5	550-D-0900	BAKER HUGHES CENTRILIFT	5,5
525-F-0900	BAKER HUGHES CENTRILIFT	5,25	550-D-1800	BAKER HUGHES CENTRILIFT	5,5
525-F-1800	BAKER HUGHES CENTRILIFT	5,25	550-D-2600	BAKER HUGHES CENTRILIFT	5,5
725-F-900	BAKER HUGHES CENTRILIFT	7,25	525-F-1300	BAKER HUGHES CENTRILIFT	5,25
725-F-1800	BAKER HUGHES CENTRILIFT	7,25	525-F-2300	BAKER HUGHES CENTRILIFT	5,25
925-F-900	BAKER HUGHES CENTRILIFT	9,25	725-F-1300	BAKER HUGHES CENTRILIFT	7,25
500-G-0900	BAKER HUGHES CENTRILIFT	5	925-F-650	BAKER HUGHES CENTRILIFT	9,25
500-G-1800	BAKER HUGHES CENTRILIFT	5	925-F-1300	BAKER HUGHES CENTRILIFT	9,25
500-G-2600	BAKER HUGHES CENTRILIFT	5	500-G-1300	BAKER HUGHES CENTRILIFT	5
700-G-1300	BAKER HUGHES CENTRILIFT	7	500-G-2300	BAKER HUGHES CENTRILIFT	5
700-G-2300	BAKER HUGHES CENTRILIFT	7	700-G-0900	BAKER HUGHES CENTRILIFT	7
900-G-0900	BAKER HUGHES CENTRILIFT	9	700-G-1800	BAKER HUGHES CENTRILIFT	7

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
900-G-1800	BAKER HUGHES CENTRILIFT	9	700-G-2600	BAKER HUGHES CENTRILIFT	7
900-G-2600	BAKER HUGHES CENTRILIFT	9	900-G-1300	BAKER HUGHES CENTRILIFT	9
1100-G-1300	BAKER HUGHES CENTRILIFT	11	900-G-2300	BAKER HUGHES CENTRILIFT	9
1100-G-2300	BAKER HUGHES CENTRILIFT	11	1100-G-0900	BAKER HUGHES CENTRILIFT	11
1400-G-0650	BAKER HUGHES CENTRILIFT	14	1100-G-1800	BAKER HUGHES CENTRILIFT	11
1400-G-1300	BAKER HUGHES CENTRILIFT	14	1100-G-2600	BAKER HUGHES CENTRILIFT	11
4-B-0	BAKER HUGHES CENTRILIFT	0,25	1400-G-0900	BAKER HUGHES CENTRILIFT	14
4-B-158	BAKER HUGHES CENTRILIFT	0,25	4-B-62	BAKER HUGHES CENTRILIFT	0,25
7-B-62	BAKER HUGHES CENTRILIFT	0,44	4-B-124	BAKER HUGHES CENTRILIFT	0,25
7-B-124	BAKER HUGHES CENTRILIFT	0,44	4-B-179	BAKER HUGHES CENTRILIFT	0,25
7-B-179	BAKER HUGHES CENTRILIFT	0,44	7-B-90	BAKER HUGHES CENTRILIFT	0,44
10-B-90	BAKER HUGHES CENTRILIFT	0,65	7-B-158	BAKER HUGHES CENTRILIFT	0,44
10-B-158	BAKER HUGHES CENTRILIFT	0,65	10-B-62	BAKER HUGHES CENTRILIFT	0,65
15-B-62	BAKER HUGHES CENTRILIFT	0,95	10-B-124	BAKER HUGHES CENTRILIFT	0,65

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
15-B-124	BAKER HUGHES CENTRILIFT	0,95	10-B-179	BAKER HUGHES CENTRILIFT	0,65
15-B-179	BAKER HUGHES CENTRILIFT	0,95	15-B-90	BAKER HUGHES CENTRILIFT	0,95
40-B-90	BAKER HUGHES CENTRILIFT	2,5	15-B-158	BAKER HUGHES CENTRILIFT	0,95
40-B-158	BAKER HUGHES CENTRILIFT	2,5	40-B-62	BAKER HUGHES CENTRILIFT	2,5
56-B-62	BAKER HUGHES CENTRILIFT	3,5	40-B-124	BAKER HUGHES CENTRILIFT	2,5
56-B-124	BAKER HUGHES CENTRILIFT	3,5	40-B-179	BAKER HUGHES CENTRILIFT	2,5
56-B-179	BAKER HUGHES CENTRILIFT	3,5	56-B-90	BAKER HUGHES CENTRILIFT	3,5
21-D-62	BAKER HUGHES CENTRILIFT	1,3	56-B-145	BAKER HUGHES CENTRILIFT	3,5
21-D-124	BAKER HUGHES CENTRILIFT	1,3	18-D-179	BAKER HUGHES CENTRILIFT	1,1
21-D-179	BAKER HUGHES CENTRILIFT	1,3	21-D-90	BAKER HUGHES CENTRILIFT	1,3
32-D-90	BAKER HUGHES CENTRILIFT	2	21-D-158	BAKER HUGHES CENTRILIFT	1,3
32-D-158	BAKER HUGHES CENTRILIFT	2	32-D-62	BAKER HUGHES CENTRILIFT	2
43-D-62	BAKER HUGHES CENTRILIFT	2,7	32-D-124	BAKER HUGHES CENTRILIFT	2
43-D-124	BAKER HUGHES CENTRILIFT	2,7	32-D-179	BAKER HUGHES CENTRILIFT	2

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
43-D-179	BAKER HUGHES CENTRILIFT	2,7	43-D-90	BAKER HUGHES CENTRILIFT	2,7
56-D-90	BAKER HUGHES CENTRILIFT	3,5	43-D-158	BAKER HUGHES CENTRILIFT	2,7
56-D-158	BAKER HUGHES CENTRILIFT	3,5	56-D-62	BAKER HUGHES CENTRILIFT	3,5
72-D-62	BAKER HUGHES CENTRILIFT	4,5	56-D-124	BAKER HUGHES CENTRILIFT	3,5
72-D-124	BAKER HUGHES CENTRILIFT	4,5	56-D-179	BAKER HUGHES CENTRILIFT	3,5
72-D-179	BAKER HUGHES CENTRILIFT	4,5	72-D-90	BAKER HUGHES CENTRILIFT	4,5
87-D-90	BAKER HUGHES CENTRILIFT	5,5	72-D-158	BAKER HUGHES CENTRILIFT	4,5
87-D-158	BAKER HUGHES CENTRILIFT	5,5	87-D-62	BAKER HUGHES CENTRILIFT	5,5
83-F-62	BAKER HUGHES CENTRILIFT	5,25	87-D-124	BAKER HUGHES CENTRILIFT	5,5
83-F-124	BAKER HUGHES CENTRILIFT	5,25	87-D-179	BAKER HUGHES CENTRILIFT	5,5
115-F-62	BAKER HUGHES CENTRILIFT	7,25	83-F-90	BAKER HUGHES CENTRILIFT	5,25
115-F-124	BAKER HUGHES CENTRILIFT	7,25	83-F-158	BAKER HUGHES CENTRILIFT	5,25
147-F-62	BAKER HUGHES CENTRILIFT	9,25	115-F-90	BAKER HUGHES CENTRILIFT	7,25
80-G-62	BAKER HUGHES CENTRILIFT	5	147-F-45	BAKER HUGHES CENTRILIFT	9,25

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
80-G-124	BAKER HUGHES CENTRILIFT	5	147-F-90	BAKER HUGHES CENTRILIFT	9,25
80-G-179	BAKER HUGHES CENTRILIFT	5	80-G-90	BAKER HUGHES CENTRILIFT	5
111-G-90	BAKER HUGHES CENTRILIFT	7	80-G-158	BAKER HUGHES CENTRILIFT	5
111-G-158	BAKER HUGHES CENTRILIFT	7	111-G-62	BAKER HUGHES CENTRILIFT	7
143-G-62	BAKER HUGHES CENTRILIFT	9	111-G-124	BAKER HUGHES CENTRILIFT	7
143-G-124	BAKER HUGHES CENTRILIFT	9	111-G-179	BAKER HUGHES CENTRILIFT	7
143-G-179	BAKER HUGHES CENTRILIFT	9	143-G-90	BAKER HUGHES CENTRILIFT	9
175-G-90	BAKER HUGHES CENTRILIFT	11	143-G-158	BAKER HUGHES CENTRILIFT	9
175-G-158	BAKER HUGHES CENTRILIFT	11	175-G-62	BAKER HUGHES CENTRILIFT	11
223-G-45	BAKER HUGHES CENTRILIFT	14	175-G-124	BAKER HUGHES CENTRILIFT	11
223-G-90	BAKER HUGHES CENTRILIFT	14	175-G-179	BAKER HUGHES CENTRILIFT	11
15TP1200	KUDU	0,19	223-G-62	BAKER HUGHES CENTRILIFT	14
30TP650	KUDU	0,34	15TP600	KUDU	0,19
30TP2000	KUDU	0,34	15TP2400	KUDU	0,19
80TP800	KUDU	1,07	30TP1300	KUDU	0,34
80TP1600	KUDU	1,07	30TP2600	KUDU	0,34
60TP650	KUDU	0,83	80TP1200	KUDU	1,07
60TP1600	KUDU	0,83	80TP2000	KUDU	1,07
60TP2600	KUDU	0,83	60TP1300	KUDU	0,83

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
100TP600	KUDU	1,36	60TP2000	KUDU	0,83
100TP1500	KUDU	1,36	60TP3300	KUDU	0,83
100TP2400	KUDU	1,36	100TP1200	KUDU	1,36
240TP900	KUDU	3,07	100TP1800	KUDU	1,36
120TP2000	KUDU	1,51	240TP600	KUDU	3,07
120TP3300	KUDU	1,51	120TP1300	KUDU	1,51
160TP1500	KUDU	2,01	120TP2600	KUDU	1,51
200TP1200	KUDU	2,44	160TP1200	KUDU	2,01
200TP1800	KUDU	2,44	160TP2200	KUDU	2,01
300TP800	KUDU	3,77	200TP1500	KUDU	2,44
300TP1600	KUDU	3,77	200TP2400	KUDU	2,44
180TP2000	KUDU	2,26	300TP1200	KUDU	3,77
225TP1600	KUDU	2,83	180TP1000	KUDU	2,26
300TP600	KUDU	3,77	180TP3000	KUDU	2,26
300TP1500	KUDU	3,77	225TP2400	KUDU	2,83
300TP2400	KUDU	3,77	300TP1800	KUDU	3,77
400TP1350	KUDU	5,03	400TP900	KUDU	5,03
500TP800	KUDU	6,29	400TP1800	KUDU	5,03
500TP1600	KUDU	6,29	500TP1200	KUDU	6,29
600TP900	KUDU	7,55	600TP600	KUDU	7,55
600TP1500	KUDU	7,55	600TP1200	KUDU	7,55
800TP600	KUDU	10,94	800TP400	KUDU	10,94
750TP800	KUDU	9,44	800TP800	KUDU	10,94
1000TP860	KUDU	13,65	750TP1200	KUDU	9,44
1000TP1450	KUDU	13,65	1000TP1150	KUDU	13,65
NTZ238*120DT14	NETZSCH	0,88	NTZ238*060DT14	NETZSCH	0,88
NTZ238*180DT14	NETZSCH	0,88	NTZ238*150DT14	NETZSCH	0,88
NTZ238*240DT14	NETZSCH	0,88	NTZ238*200DT14	NETZSCH	0,88
NTZ278*090DT16	NETZSCH	1	NTZ278*060DT16	NETZSCH	1
NTZ278*150DT16	NETZSCH	1	NTZ278*120DT16	NETZSCH	1
NTZ278*200DT16	NETZSCH	1	NTZ278*180DT16	NETZSCH	1

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)
NTZ278*09 0DT20	NETZSCH	1,26	NTZ278*24 0DT16	NETZSCH	1
NTZ278*15 0DT20	NETZSCH	1,26	NTZ278*12 0DT20	NETZSCH	1,26
NTZ278*20 0DT20	NETZSCH	1,26	NTZ278*18 0DT20	NETZSCH	1,26
NTZ278*06 0DT25	NETZSCH	1,57	NTZ278*24 0DT20	NETZSCH	1,26
NTZ278*10 0DT25	NETZSCH	1,57	NTZ278*09 0DT25	NETZSCH	1,57
NTZ278*15 0DT25	NETZSCH	1,57	NTZ278*12 0DT25	NETZSCH	1,57
NTZ278*20 0DT25	NETZSCH	1,57	NTZ278*18 0DT25	NETZSCH	1,57
NTZ278*04 8DT32	NETZSCH	2	NTZ278*24 0DT25	NETZSCH	1,57
NTZ278*09 0DT32	NETZSCH	2	NTZ278*06 0DT32	NETZSCH	2
NTZ278*15 0DT32	NETZSCH	2	NTZ278*12 0DT32	NETZSCH	2
NTZ278*20 0DT32	NETZSCH	2	NTZ278*18 0DT32	NETZSCH	2
NTZ350*06 0DT33	NETZSCH	2,08	NTZ278*24 0DT32	NETZSCH	2
NTZ350*10 0DT33	NETZSCH	2,08	NTZ350*09 0DT33	NETZSCH	2,08
NTZ350*15 0DT33	NETZSCH	2,08	NTZ350*12 0DT33	NETZSCH	2,08
NTZ350*20 0DT33	NETZSCH	2,08	NTZ350*18 0DT33	NETZSCH	2,08
NTZ350*30 0DT33	NETZSCH	2,08	NTZ350*24 0DT33	NETZSCH	2,08
NTZ350*12 0DT40	NETZSCH	2,5	NTZ350*10 0DT40	NETZSCH	2,5
NTZ350*20 0DT40	NETZSCH	2,5	NTZ350*15 0DT40	NETZSCH	2,5
NTU350*16 0DT40	NETZSCH	2,5	NTU350*12 0DT40	NETZSCH	2,5
NTU350*32 0DT40	NETZSCH	2,5	NTU350*24 0DT40	NETZSCH	2,5
NTZ400*09 0DT50	NETZSCH	3,15	NTZ400*06 0DT50	NETZSCH	3,15
NTZ400*12 0DT50	NETZSCH	3,15	NTZ400*10 0DT50	NETZSCH	3,15

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm) 4
NTZ400*18 0DT50	NETZSCH	3,15	NTZ400*15 0DT50	NETZSCH	3,15
NTZ400*24 0DT50	NETZSCH	3,15	NTZ400*20 0DT50	NETZSCH	3,15
NTZ400*09 0DT66	NETZSCH	4,15	NTZ400*06 0DT66	NETZSCH	4,15
NTZ400*12 0DT66	NETZSCH	4,15	NTZ400*10 0DT66	NETZSCH	4,15
NTZ400*18 0DT66	NETZSCH	4,15	NTZ400*15 0DT66	NETZSCH	4,15
NTZ400*24 0DT66	NETZSCH	4,15	NTZ400*20 0DT66	NETZSCH	4,15
NTZ400*09 0DT83	NETZSCH	5,22	NTZ400*08 0DT83	NETZSCH	5,22
NTZ400*12 0DT83	NETZSCH	5,22	NTZ400*10 0DT83	NETZSCH	5,22
NTZ400*18 0DT83	NETZSCH	5,22	NTZ400*15 0DT83	NETZSCH	5,22
NTU400*07 0DT83	NETZSCH	5,22	NTZ400*20 0DT83	NETZSCH	5,22
NTU400*13 0DT83	NETZSCH	5,22	NTU400*10 0DT83	NETZSCH	5,22
NTU400*17 0DT83	NETZSCH	5,22	NTU400*20 0DT83	NETZSCH	5,22
NTZ400*06 0DT110	NETZSCH	6,92	NTZ400*09 0DT110	NETZSCH	6,92
NTZ400*12 0DT110	NETZSCH	6,92	NTZ400*15 0DT110	NETZSCH	6,92
NTZ400*06 0DT142	NETZSCH	8,93	NTZ400*09 0DT142	NETZSCH	8,93
NTZ400*12 0DT142	NETZSCH	8,93	NTZ450*09 0DT74	NETZSCH	4,65
NTZ450*12 0DT74	NETZSCH	4,65	NTZ450*15 0DT74	NETZSCH	4,65
NTZ450*18 0DT74	NETZSCH	4,65	NTZ450*20 0DT74	NETZSCH	4,65
NTZ450*24 0DT74	NETZSCH	4,65	NTZ450*09 0DT150	NETZSCH	9,43
NTZ450*12 0DT150	NETZSCH	9,43	NTZ450*15 0DT150	NETZSCH	9,43
NTZ500*09 0DT138	NETZSCH	8,68	NTZ500*12 0DT138	NETZSCH	8,68
NTZ500*15 0DT138	NETZSCH	8,68	NTZ500*18 0DT138	NETZSCH	8,68

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)
NTZ500*20 0DT138	NETZSCH	8,68	NTZ500*10 0DT170	NETZSCH	10,69
NTZ500*12 0DT170	NETZSCH	10,69	NTZ500*15 0DT170	NETZSCH	10,69
NTZ500*18 0DT170	NETZSCH	10,69	NTZ500*09 0DT226	NETZSCH	14,21
NTZ500*12 0DT226	NETZSCH	14,21	870-D-88	NETZSCH	0,88
1800-D-88	NETZSCH	0,88	2200-D-88	NETZSCH	0,88
2600-D-88	NETZSCH	0,88	2900-D-88	NETZSCH	0,88
3500-D-88	NETZSCH	0,88	870-D-100	NETZSCH	1
1300-D-100	NETZSCH	1	1800D- D100	NETZSCH	1
2200-D-100	NETZSCH	1	2600-D-100	NETZSCH	1
2900-D-100	NETZSCH	1	3500-D-100	NETZSCH	1
1300-D-126	NETZSCH	1,26	1800-D-126	NETZSCH	1,26
2200-D-126	NETZSCH	1,26	2600-D-126	NETZSCH	1,26
2900-D-126	NETZSCH	1,26	3500-D-126	NETZSCH	1,26
870-D-157	NETZSCH	1,57	1300-D-157	NETZSCH	1,57
1500-D-157	NETZSCH	1,57	1800-D-157	NETZSCH	1,57
2200-D-157	NETZSCH	1,57	2600-D-157	NETZSCH	1,57
2900-D-157	NETZSCH	1,57	3500-D-157	NETZSCH	1,57
700-D-200	NETZSCH	2	870-D-200	NETZSCH	2
1300-D-200	NETZSCH	2	1800-D-200	NETZSCH	2
2200-D-200	NETZSCH	2	2600-D-200	NETZSCH	2
2900-D-200	NETZSCH	2	3500-D-200	NETZSCH	2
870-D-200	NETZSCH	2,08	1300-D-208	NETZSCH	2,08
1500-D-208	NETZSCH	2,08	1800-D-208	NETZSCH	2,08
2200-D-208	NETZSCH	2,08	2600-D-208	NETZSCH	2,08
2400-D-208	NETZSCH	2,08	3500-D-208	NETZSCH	2,08
4400-D-208	NETZSCH	2,08	1500-D-250	NETZSCH	2,5
1800-D-250	NETZSCH	2,5	2200-D-250	NETZSCH	2,5
2900-D-250	NETZSCH	2,5	1800-D-250	NETZSCH	2,5
2320-D-250	NETZSCH	2,5	3500-D-250	NETZSCH	2,5
4640-D-250	NETZSCH	2,5	870-D-315	NETZSCH	3,15
1300-D-315	NETZSCH	3,15	1500-D-315	NETZSCH	3,15
1800-D-315	NETZSCH	3,15	2200-D-315	NETZSCH	3,15
2600-D-315	NETZSCH	3,15	2900-D-315	NETZSCH	3,15

MODELO	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)	MODELO.	MARCA	DESPLAZA- MIENTO (bbl/dia*rpm)
3500-D-315	NETZSCH	3,15	870-D-415	NETZSCH	4,15
1300-D-415	NETZSCH	4,15	1500-D-415	NETZSCH	4,15
1800-D-415	NETZSCH	4,15	2200-D-415	NETZSCH	4,15
2600-D-415	NETZSCH	4,15	2900-D-415	NETZSCH	4,15
3500-D-415	NETZSCH	4,15	1160-D-522	NETZSCH	5,22
1300-D-522	NETZSCH	5,22	1500-D-522	NETZSCH	5,22
1800-D-522	NETZSCH	5,22	2200-D-522	NETZSCH	5,22
2600-D-522	NETZSCH	5,22	2900-D-522	NETZSCH	5,22
1010-D-522	NETZSCH	5,22	1300-D-692	NETZSCH	6,92
1900-D-522	NETZSCH	5,22	2200-D-692	NETZSCH	6,92
2500-D-522	NETZSCH	5,22	1300-D-893	NETZSCH	8,93
870-D-692	NETZSCH	6,92	1300-D-465	NETZSCH	4,65
1800-D-692	NETZSCH	6,92	2200-D-465	NETZSCH	4,65
870-D-893	NETZSCH	8,93	2900-D-465	NETZSCH	4,65
1800-D-893	NETZSCH	8,93	1300-D-943	NETZSCH	9,43
1800-D-465	NETZSCH	4,65	2200-D-943	NETZSCH	9,43
2600-D-465	NETZSCH	4,65	1800-D-868	NETZSCH	8,68
3500-D-465	NETZSCH	4,65	2600-D-868	NETZSCH	8,68
1800-D-943	NETZSCH	9,43	1500-D-1069	NETZSCH	10,69
1300-D-868	NETZSCH	8,68	2200-D-1069	NETZSCH	10,69
2200-D-868	NETZSCH	8,68	1800-D-1069	NETZSCH	10,69
2900-D-868	NETZSCH	8,68	2600-D-1069	NETZSCH	10,69
1800-D-1421	NETZSCH	14,21	1300-D-1421	NETZSCH	14,21
NTZ550*13 5ST200	NETZSCH	12,58	NTZ550*12 0SIT78	NETZSCH	4,91

Anexo 5: Resultados de la regresión lineal multivariable (Set de datos reales caracterizados)

Anexo 5. Resultados de la regresión lineal multivariable

NOMBRE	TIPO	AMBIENTE	YACIMIENTO	T (s)	TOTAL DE DATOS ANALIZADOS	r²	ERROR RELATIVO PROMEDIO (%)	ERROR ABSOLUTO PROMEDIO (bpd)
VD01	Entrena	DELTAICO	B2	0.48	63526.00	0.93	22.45	115.6858
VD01	Prueba	DELTAICO	B2	0.48	63526.00	0.93	33.43	225.0416
VD01	Total	DELTAICO	B2	0.48	63526.00	0.93	25.74	148.4981
VD02	Entrena	FLUVIAL	D3	0.28	63526.00	0.96	20.29	93.2467
VD02	Prueba	FLUVIAL	D3	0.28	63526.00	0.96	56.79	133.8709
VD02	Total	FLUVIAL	D3	0.28	63526.00	0.96	31.24	105.4360
VD03	Entrena	DELTAICO	B2	0.14	55260.00	0.90	20.11	122.8800
VD03	Prueba	DELTAICO	B2	0.14	55260.00	0.90	26.39	150.3681
VD03	Total	DELTAICO	B2	0.14	55260.00	0.90	21.99	131.1209
VD04	Entrena	FLUVIAL	D3	0.14	63549.00	0.92	16.13	128.5673
VD04	Prueba	FLUVIAL	D3	0.14	63549.00	0.92	36.35	157.9527
VD04	Total	FLUVIAL	D3	0.14	63549.00	0.92	22.19	137.3770
VD05	Entrena	DELTAICO	B2	0.14	66288.00	0.91	23.53	119.4708
VD05	Prueba	DELTAICO	B2	0.14	66288.00	0.91	27.75	115.1285
VD05	Total	DELTAICO	B2	0.14	66288.00	0.91	24.80	118.1679
VD06	Entrena	FLUVIAL	D3	0.11	55240.00	0.94	21.50	89.4621
VD06	Prueba	FLUVIAL	D3	0.11	55240.00	0.94	44.39	139.5756
VD06	Total	FLUVIAL	D3	0.11	55240.00	0.94	28.36	104.4986
VD07	Entrena	DELTAICO	B2	0.10	58023.00	0.94	16.53	95.5490
VD07	Prueba	DELTAICO	B2	0.10	58023.00	0.94	37.78	186.1709
VD07	Total	DELTAICO	B2	0.10	58023.00	0.94	22.90	122.7172
VD08	Entrena	FLUVIAL	D3	0.13	66336.00	0.95	24.33	54.9132
VD08	Prueba	FLUVIAL	D3	0.13	66336.00	0.95	19.04	46.1631
VD08	Total	FLUVIAL	D3	0.13	66336.00	0.95	22.74	52.2922
VD09	Entrena	DELTAICO	B2	0.11	57981.00	0.91	19.67	119.8482
VD09	Prueba	DELTAICO	B2	0.11	57981.00	0.91	30.42	100.9957
VD09	Total	DELTAICO	B2	0.11	57981.00	0.91	22.90	114.1867
VD10	Entrena	DELTAICO	C2	0.09	60764.00	0.93	9.24	22.7169
VD10	Prueba	DELTAICO	C2	0.09	60764.00	0.93	8.53	32.9429
VD10	Total	DELTAICO	C2	0.09	60764.00	0.93	9.03	25.7852

Anexo 5. Resultados de la regresión lineal multivariable (Continuación)

NOMBRE	TIPO	AMBIENTE	YACIMIENTO	T (s)	TOTAL DE DATOS ANALIZADOS	r2	ERROR RELATIVO PROMEDIO (%)	ERROR ABSOLUTO PROMEDIO (bpd)
VD11	Entrena	FLUVIAL	D3	0.14	60742.00	0.94	19.93	58.4030
VD11	Prueba	FLUVIAL	D3	0.14	60742.00	0.94	20.78	58.4775
VD11	Total	FLUVIAL	D3	0.14	60742.00	0.94	20.18	58.4254
VD12	Entrena	DELTAICO	B2	0.14	66288.00	0.90	15.21	79.1166
VD12	Prueba	DELTAICO	B2	0.14	66288.00	0.90	18.82	93.3444
VD12	Total	DELTAICO	B2	0.14	66288.00	0.90	16.30	83.3857
VD13	Entrena	FLUVIAL	D3	0.16	63480.00	0.94	183.37	117.5784
VD13	Prueba	FLUVIAL	D3	0.16	63480.00	0.94	62.36	164.7047
VD13	Total	FLUVIAL	D3	0.16	63480.00	0.94	147.00	131.7425
VD14	Entrena	DELTAICO	B2	0.12	63549.00	0.92	16.11	70.0323
VD14	Prueba	DELTAICO	B2	0.12	63549.00	0.92	20.37	89.7976
VD14	Total	DELTAICO	B2	0.12	63549.00	0.92	17.39	75.9579
VD15	Entrena	DELTAICO	B2	0.09	58023.00	0.93	21.52	61.3679
VD15	Prueba	DELTAICO	B2	0.09	58023.00	0.93	34.56	58.3304
VD15	Total	DELTAICO	B2	0.09	58023.00	0.93	25.42	60.4573
VD16	Entrena	DELTAICO	B2	0.17	60764.00	0.93	20.12	85.6926
VD16	Prueba	DELTAICO	B2	0.17	60764.00	0.93	47.10	134.7252
VD16	Total	DELTAICO	B2	0.17	60764.00	0.93	28.21	100.4049
VD17	Entrena	DELTAICO	B2	0.12	55240.00	0.93	17.58	105.7734
VD17	Prueba	DELTAICO	B2	0.12	55240.00	0.93	31.35	155.1233
VD17	Total	DELTAICO	B2	0.12	55240.00	0.93	21.71	120.5809
VD18	Entrena	FLUVIAL	D3	0.33	74574.00	0.92	20.84	155.6457
VD18	Prueba	FLUVIAL	D3	0.33	74574.00	0.92	90.30	557.1173
VD18	Total	FLUVIAL	D3	0.33	74574.00	0.92	41.68	276.1075
VD19	Entrena	DELTAICO	B2	0.22	66288.00	0.90	15.69	116.8950
VD19	Prueba	DELTAICO	B2	0.22	66288.00	0.90	25.19	143.6238
VD19	Total	DELTAICO	B2	0.22	66288.00	0.90	18.54	124.9150
VD20	Entrena	FLUVIAL	D3	0.27	71812.00	0.90	13.01	23.9379
VD20	Prueba	FLUVIAL	D3	0.27	71812.00	0.90	21.39	27.0007
VD20	Total	FLUVIAL	D3	0.27	71812.00	0.90	15.52	24.8569
VD21	Entrena	DELTAICO	B2	0.16	60764.00	0.94	14.42	101.9508
VD21	Prueba	DELTAICO	B2	0.16	60764.00	0.94	24.76	154.7916
VD21	Total	DELTAICO	B2	0.16	60764.00	0.94	17.53	117.8057

Anexo 5. Resultados de la regresión lineal multivariable (Continuación)

NOMBRE	TIPO	AMBIENTE	YACIMIENTO	T (s)	TOTAL DE DATOS ANALIZADOS	r2	ERROR RELATIVO PROMEDIO (%)	ERROR ABSOLUTO PROMEDIO (bpd)
VD22	Entrena	FLUVIAL	E1	0.35	52478.00	0.98	4.02	23.8389
VD22	Prueba	FLUVIAL	E1	0.35	52478.00	0.98	5.07	33.7743
VD22	Total	FLUVIAL	E1	0.35	52478.00	0.98	4.34	26.8201
VD23	Entrena	FLUVIAL	D3	0.25	63526.00	0.94	13.29	13.1052
VD23	Prueba	FLUVIAL	D3	0.25	63526.00	0.94	10.75	17.4326
VD23	Total	FLUVIAL	D3	0.25	63526.00	0.94	12.53	14.4036
VD24	Entrena	FLUVIAL	D3	0.24	63526.00	0.92	22.83	106.5370
VD24	Prueba	FLUVIAL	D3	0.24	63526.00	0.92	26.03	126.8639
VD24	Total	FLUVIAL	D3	0.24	63526.00	0.92	23.79	112.6361
VD25	Entrena	FLUVIAL	E1	0.11	28622.00	0.96	16.52	31.2412
VD25	Prueba	FLUVIAL	E1	0.11	28622.00	0.96	300.90	51.9279
VD25	Total	FLUVIAL	E1	0.11	28622.00	0.96	101.88	37.4506
VD26	Entrena	FLUVIAL	E1	0.08	27006.00	0.94	31.81	23.8176
VD26	Prueba	FLUVIAL	E1	0.08	27006.00	0.94	17.11	44.2167
VD26	Total	FLUVIAL	E1	0.08	27006.00	0.94	27.40	29.9362
VD27	Entrena	FLUVIAL	D3	0.12	30038.00	0.95	20.80	32.0985
VD27	Prueba	FLUVIAL	D3	0.12	30038.00	0.95	45.81	87.6505
VD27	Total	FLUVIAL	D3	0.12	30038.00	0.95	28.32	48.7909
VD28	Entrena	FLUVIAL	E1	0.12	30199.00	0.94	10.96	17.4839
VD28	Prueba	FLUVIAL	E1	0.12	30199.00	0.94	19.69	17.8044
VD28	Total	FLUVIAL	E1	0.12	30199.00	0.94	13.57	17.5797
VD29	Entrena	FLUVIAL	D3	0.26	27447.00	0.93	16.75	102.2351
VD29	Prueba	FLUVIAL	D3	0.26	27447.00	0.93	21.45	133.1947
VD29	Total	FLUVIAL	D3	0.26	27447.00	0.93	18.16	111.5213
VD30	Entrena	FLUVIAL	D3	0.14	28644.00	0.93	11.62	72.7405
VD30	Prueba	FLUVIAL	D3	0.14	28644.00	0.93	19.97	73.1201
VD30	Total	FLUVIAL	D3	0.14	28644.00	0.93	14.12	72.8544
VD31	Entrena	FLUVIAL	E1	0.07	27594.00	0.95	17.02	49.3033
VD31	Prueba	FLUVIAL	E1	0.07	27594.00	0.95	35.15	63.8077
VD31	Total	FLUVIAL	E1	0.07	27594.00	0.95	22.45	53.6484
VD32	Entrena	FLUVIAL	E1	0.07	25740.00	0.96	8.15	20.3924
VD32	Prueba	FLUVIAL	E1	0.07	25740.00	0.96	6.73	15.8494

Anexo 5. Resultados de la regresión lineal multivariable (Continuación)

NOMBRE	TIPO	AMBIENTE	YACIMIENTO	T (s)	TOTAL DE DATOS ANALIZADOS	r2	ERROR RELATIVO PROMEDIO (%)	ERROR ABSOLUTO PROMEDIO (bpd)
VD32	Total	FLUVIAL	E1	0.07	25740.00	0.96	7.73	19.0323
VD33	Entrena	FLUVIAL	D3	0.08	27930.00	0.97	49.52	64.0575
VD33	Prueba	FLUVIAL	D3	0.08	27930.00	0.97	42.66	124.1693
VD33	Total	FLUVIAL	D3	0.08	27930.00	0.97	47.46	82.0910
VD34	Entrena	DELTAICO	C2	0.13	30613.00	0.89	13.23	87.9433
VD34	Prueba	DELTAICO	C2	0.13	30613.00	0.89	15.15	81.9610
VD34	Total	DELTAICO	C2	0.13	30613.00	0.89	13.81	86.1495
VD35	Entrena	DELTAICO	C2	0.08	27930.00	0.95	9.43	37.5631
VD35	Prueba	DELTAICO	C2	0.08	27930.00	0.95	31.29	72.6584
VD35	Total	DELTAICO	C2	0.08	27930.00	0.95	16.00	48.1046
VD36	Entrena	DELTAICO	C2	0.11	28644.00	0.95	10.23	48.5447
VD36	Prueba	DELTAICO	C2	0.11	28644.00	0.95	22.14	81.3107
VD36	Total	DELTAICO	C2	0.11	28644.00	0.95	13.80	58.3621
VD37	Entrena	DELTAICO	C2	0.34	24434.00	0.95	12.22	5.0793
VD37	Prueba	DELTAICO	C2	0.34	24434.00	0.95	402.69	56.8081
VD37	Total	DELTAICO	C2	0.34	24434.00	0.95	129.34	20.5951
VD38	Entrena	DELTAICO	C2	0.27	26040.00	0.99	2.25	3.8719
VD38	Prueba	DELTAICO	C2	0.27	26040.00	0.99	0.00	0.0000
VD38	Total	DELTAICO	C2	0.27	26040.00	0.99	1.57	2.7118
VD39	Entrena	DELTAICO	C2	0.05	25820.00	0.98	6.75	6.5788
VD39	Prueba	DELTAICO	C2	0.05	25820.00	0.98	81.40	14.9757
VD39	Total	DELTAICO	C2	0.05	25820.00	0.98	29.17	9.1001
VD40	Entrena	DELTAICO	C2	0.13	28402.00	0.97	11.71	45.6302
VD40	Prueba	DELTAICO	C2	0.13	28402.00	0.97	6.39	21.8882
VD40	Total	DELTAICO	C2	0.13	28402.00	0.97	10.11	38.5012

Anexo 6: Resultados de la regresión polinomial (Set de datos reales caracterizados)

Anexo 6. Resultados de la regresión polinomial

GRADO	NOMBRE	TIPO	YACI-MIEN-TO	NUME-RO DE COEFI-CIEN-TES	r²	TIEMPO (s)	ERROR RELATIVO PROMEDIO (%)	ERROR ABSOLUTO PROMEDIO (bpd)
Degree 2	VD01	Entrena	B2	253	0.9873	24.390	19.051	52.1099
Degree 2	VD01	Prueba	B2	253	0.9873	24.390	67.632	229.9179
Degree 2	VD01	Total	B2	253	0.9873	24.390	33.628	105.4613
Degree 3	VD01	Entrena	B2	2024	0.9972	63.928	3.590	20.7063
Degree 3	VD01	Prueba	B2	2024	0.9972	63.928	122.633	442.0326
Degree 3	VD01	Total	B2	2024	0.9972	63.928	39.309	147.1255
Degree 4	VD01	Entrena	B2	12650	0.9978	272.298	4.298	18.5993
Degree 4	VD01	Prueba	B2	12650	0.9978	272.298	217.908	1198.421
Degree 4	VD01	Total	B2	12650	0.9978	272.298	68.392	372.6056
Degree 2	VD02	Entrena	D3	253	0.9963	28.128	4.257	26.0281
Degree 2	VD02	Prueba	D3	253	0.9963	28.128	30.958	93.8282
Degree 2	VD02	Total	D3	253	0.9963	28.128	12.268	46.3716
Degree 3	VD02	Entrena	D3	2024	0.9992	75.780	2.085	12.0736
Degree 3	VD02	Prueba	D3	2024	0.9992	75.780	162.074	271.1319
Degree 3	VD02	Total	D3	2024	0.9992	75.780	50.089	89.8042
Degree 4	VD02	Entrena	D3	12650	0.9996	283.431	1.442	8.1505
Degree 4	VD02	Prueba	D3	12650	0.9996	283.431	1055.955	752.3575
Degree 4	VD02	Total	D3	12650	0.9996	283.431	317.849	231.4503
Degree 2	VD03	Entrena	B2	190	0.9902	27.057	6.114	36.4674
Degree 2	VD03	Prueba	B2	190	0.9902	27.057	23.418	133.8313
Degree 2	VD03	Total	B2	190	0.9902	27.057	11.302	65.6568
Degree 3	VD03	Entrena	B2	1330	0.9988	62.337	2.173	13.589
Degree 3	VD03	Prueba	B2	1330	0.9988	62.337	51.520	249.1451
Degree 3	VD03	Total	B2	1330	0.9988	62.337	16.967	84.2081
Degree 4	VD03	Entrena	B2	7315	0.9994	195.557	1.595	10.0082
Degree 4	VD03	Prueba	B2	7315	0.9994	195.557	87.545	390.0538
Degree 4	VD03	Total	B2	7315	0.9994	195.557	27.363	123.9448
Degree 2	VD04	Entrena	D3	253	0.9919	26.764	4.115	38.4815
Degree 2	VD04	Prueba	D3	253	0.9919	26.764	31.494	178.1543
Degree 2	VD04	Total	D3	253	0.9919	26.764	12.323	80.355
Degree 3	VD04	Entrena	D3	2024	0.9984	71.868	2.015	17.7028

Anexo 6. Resultados de la regresión polinomial (Continuación)

GRADO	NOMBRE	TIPO	YACI-MIENTO	NUMERO DE COEFICIENTES	r ²	TIEMPO (s)	ERROR RELATIVO PROMEDIO (%)	ERROR ABSOLUTO PROMEDIO (bpd)
Degree 3	VD04	Prueba	D3	2024	0.9984	71.868	95.203	472.7259
Degree 3	VD04	Total	D3	2024	0.9984	71.868	29.953	154.1175
Degree 4	VD04	Entrena	D3	12650	0.9991	265.844	1.572	13.6205
Degree 4	VD04	Prueba	D3	12650	0.9991	265.844	114.844	820.4685
Degree 4	VD04	Total	D3	12650	0.9991	265.844	35.530	255.5113
Degree 2	VD05	Entrena	B2	276	0.9942	27.092	5.568	29.0783
Degree 2	VD05	Prueba	B2	276	0.9942	27.092	31.236	112.2827
Degree 2	VD05	Total	B2	276	0.9942	27.092	13.270	54.0438
Degree 3	VD05	Entrena	B2	2300	0.9992	70.166	1.610	10.2109
Degree 3	VD05	Prueba	B2	2300	0.9992	70.166	53.598	267.9374
Degree 3	VD05	Total	B2	2300	0.9992	70.166	17.209	87.5419
Degree 4	VD05	Entrena	B2	14950	0.9994	298.187	1.266	8.3549
Degree 4	VD05	Prueba	B2	14950	0.9994	298.187	55.288	257.2579
Degree 4	VD05	Total	B2	14950	0.9994	298.187	17.475	83.0384
Degree 2	VD06	Entrena	D3	190	0.9942	22.496	5.093	27.2434
Degree 2	VD06	Prueba	D3	190	0.9942	22.496	30.232	136.5973
Degree 2	VD06	Total	D3	190	0.9942	22.496	12.636	60.0551
Degree 3	VD06	Entrena	D3	1330	0.9989	54.352	2.547	12.0704
Degree 3	VD06	Prueba	D3	1330	0.9989	54.352	63.203	258.667
Degree 3	VD06	Total	D3	1330	0.9989	54.352	20.747	86.0618
Degree 4	VD06	Entrena	D3	7315	0.9996	169.881	1.404	7.3221
Degree 4	VD06	Prueba	D3	7315	0.9996	169.881	262.195	650.9955
Degree 4	VD06	Total	D3	7315	0.9996	169.881	79.654	200.4567
Degree 2	VD07	Entrena	B2	210	0.9924	20.595	4.432	30.474
Degree 2	VD07	Prueba	B2	210	0.9924	20.595	24.635	205.0656
Degree 2	VD07	Total	B2	210	0.9924	20.595	10.489	82.8161
Degree 3	VD07	Entrena	B2	1540	0.9906	50.938	6.973	38.0259
Degree 3	VD07	Prueba	B2	1540	0.9906	50.938	85.655	730.3596
Degree 3	VD07	Total	B2	1540	0.9906	50.938	30.561	245.5856
Degree 4	VD07	Entrena	B2	8855	0.9988	183.682	1.989	13.036
Degree 4	VD07	Prueba	B2	8855	0.9988	183.682	82.233	856.344
Degree 4	VD07	Total	B2	8855	0.9988	183.682	26.046	265.8574
Degree 2	VD08	Entrena	D3	276	0.9920	25.483	7.371	19.1191
Degree 2	VD08	Prueba	D3	276	0.9920	25.483	25.492	49.9675

Anexo 6. Resultados de la regresión polinomial (Continuación)

GRADO	NOMBRE	TIPO	YACI- MIEN- TO	NUME- RO DE COEFI- CIEN- TES	r ²	TIEMPO (s)	ERROR RELATIVO PROMEDIO (%)	ERROR ABSOLUTO PROMEDIO (bpd)
Degree 2	VD08	Total	D3	276	0.9920	25.483	12.799	28.3595
Degree 3	VD08	Entrena	D3	2300	0.9989	74.452	3.186	8.0787
Degree 3	VD08	Prueba	D3	2300	0.9989	74.452	70.907	201.9332
Degree 3	VD08	Total	D3	2300	0.9989	74.452	23.471	66.1466
Degree 4	VD08	Entrena	D3	14950	0.9998	313.464	1.604	3.8326
Degree 4	VD08	Prueba	D3	14950	0.9998	313.464	369.113	582.3849
Degree 4	VD08	Total	D3	14950	0.9998	313.464	111.689	177.1344
Degree 2	VD09	Entrena	B2	210	0.9876	21.274	8.537	41.2161
Degree 2	VD09	Prueba	B2	210	0.9876	21.274	64.245	177.9622
Degree 2	VD09	Total	B2	210	0.9876	21.274	25.266	82.2815
Degree 3	VD09	Entrena	B2	1540	0.9943	51.407	6.137	30.6979
Degree 3	VD09	Prueba	B2	1540	0.9943	51.407	198.981	512.8008
Degree 3	VD09	Total	B2	1540	0.9943	51.407	64.049	175.4753
Degree 4	VD09	Entrena	B2	8855	0.9983	183.451	2.594	16.1624
Degree 4	VD09	Prueba	B2	8855	0.9983	183.451	173.387	399.3684
Degree 4	VD09	Total	B2	8855	0.9983	183.451	53.884	131.2407
Degree 2	VD10	Entrena	C2	231	0.9976	24.046	1.818	5.0418
Degree 2	VD10	Prueba	C2	231	0.9976	24.046	7.960	20.7598
Degree 2	VD10	Total	C2	231	0.9976	24.046	3.661	9.758
Degree 3	VD10	Entrena	C2	1771	0.9998	60.200	0.557	1.4605
Degree 3	VD10	Prueba	C2	1771	0.9998	60.200	180.186	373.9042
Degree 3	VD10	Total	C2	1771	0.9998	60.200	54.455	113.2125
Degree 4	VD10	Entrena	C2	10626	1.0000	222.231	0.164	0.4445
Degree 4	VD10	Prueba	C2	10626	1.0000	222.231	232.651	584.0721
Degree 4	VD10	Total	C2	10626	1.0000	222.231	69.922	175.5624
Degree 2	VD11	Entrena	D3	231	0.9934	21.041	5.164	17.2661
Degree 2	VD11	Prueba	D3	231	0.9934	21.041	31.535	85.2487
Degree 2	VD11	Total	D3	231	0.9934	21.041	13.084	37.6816
Degree 3	VD11	Entrena	D3	1771	0.9983	54.147	3.398	9.5767
Degree 3	VD11	Prueba	D3	1771	0.9983	54.147	94.386	211.7253
Degree 3	VD11	Total	D3	1771	0.9983	54.147	30.722	70.2828
Degree 4	VD11	Entrena	D3	10626	0.9994	220.505	1.615	4.9892
Degree 4	VD11	Prueba	D3	10626	0.9994	220.505	125.567	403.5308
Degree 4	VD11	Total	D3	10626	0.9994	220.505	38.839	124.6729

Anexo 6. Resultados de la regresión polinomial (Continuación)

GRADO	NOMBRE	TIPO	YACI-MIENTO	NUMERO DE COEFICIENTES	r ²	TIEMPO (s)	ERROR RELATIVO PROMEDIO (%)	ERROR ABSOLUTO PROMEDIO (bpd)
Degree 2	VD12	Entrena	B2	276	0.9924	25.014	3.354	20.1199
Degree 2	VD12	Prueba	B2	276	0.9924	25.014	16.300	92.9985
Degree 2	VD12	Total	B2	276	0.9924	25.014	7.239	41.9871
Degree 3	VD12	Entrena	B2	2300	0.9987	68.383	1.699	8.5855
Degree 3	VD12	Prueba	B2	2300	0.9987	68.383	70.377	410.6532
Degree 3	VD12	Total	B2	2300	0.9987	68.383	22.306	129.2262
Degree 4	VD12	Entrena	B2	14950	0.9992	312.113	1.152	6.4292
Degree 4	VD12	Prueba	B2	14950	0.9992	312.113	136.216	825.4184
Degree 4	VD12	Total	B2	14950	0.9992	312.113	41.678	252.1675
Degree 2	VD13	Entrena	D3	253	0.9898	25.860	7.696	43.4198
Degree 2	VD13	Prueba	D3	253	0.9898	25.860	29.138	139.7574
Degree 2	VD13	Total	D3	253	0.9898	25.860	14.140	72.3748
Degree 3	VD13	Entrena	D3	2024	0.9980	67.959	8.922	22.2755
Degree 3	VD13	Prueba	D3	2024	0.9980	67.959	137.972	630.63
Degree 3	VD13	Total	D3	2024	0.9980	67.959	47.709	205.121
Degree 4	VD13	Entrena	D3	12650	0.9997	270.871	2.093	7.4783
Degree 4	VD13	Prueba	D3	12650	0.9997	270.871	97.047	607.4074
Degree 4	VD13	Total	D3	12650	0.9997	270.871	30.632	187.7915
Degree 2	VD14	Entrena	B2	253	0.9894	22.252	5.402	24.4822
Degree 2	VD14	Prueba	B2	253	0.9894	22.252	27.179	116.4231
Degree 2	VD14	Total	B2	253	0.9894	22.252	11.931	52.0458
Degree 3	VD14	Entrena	B2	2024	0.9986	59.099	1.429	7.9211
Degree 3	VD14	Prueba	B2	2024	0.9986	59.099	80.207	351.0391
Degree 3	VD14	Total	B2	2024	0.9986	59.099	25.047	110.7869
Degree 4	VD14	Entrena	B2	12650	0.9989	258.837	1.281	6.873
Degree 4	VD14	Prueba	B2	12650	0.9989	258.837	427.432	1822.6949
Degree 4	VD14	Total	B2	12650	0.9989	258.837	129.040	551.2514
Degree 2	VD15	Entrena	B2	210	0.9843	21.433	9.753	29.1738
Degree 2	VD15	Prueba	B2	210	0.9843	21.433	35.497	60.5047
Degree 2	VD15	Total	B2	210	0.9843	21.433	17.471	38.5667
Degree 3	VD15	Entrena	B2	1540	0.9974	51.901	5.450	12.6013
Degree 3	VD15	Prueba	B2	1540	0.9974	51.901	73.544	161.5722
Degree 3	VD15	Total	B2	1540	0.9974	51.901	25.864	57.2624
Degree 4	VD15	Entrena	B2	8855	0.9991	189.391	2.840	7.5428

Anexo 6. Resultados de la regresión polinomial (Continuación)

GRADO	NOMBRE	TIPO	YACI- MIEN- TO	NUME- RO DE COEFI- CIEN- TES	r ²	TIEMPO (s)	ERROR RELATIVO PROMEDIO (%)	ERROR ABSOLUTO PROMEDIO (bpd)
Degree 4	VD15	Prueba	B2	8855	0.9991	189.391	207.335	298.4542
Degree 4	VD15	Total	B2	8855	0.9991	189.391	64.147	94.7572
Degree 2	VD16	Entrena	B2	210	0.9836	21.254	9.146	40.9103
Degree 2	VD16	Prueba	B2	210	0.9836	21.254	310.138	444.2013
Degree 2	VD16	Total	B2	210	0.9836	21.254	99.459	161.9181
Degree 3	VD16	Entrena	B2	1540	0.9983	50.889	3.183	13.3329
Degree 3	VD16	Prueba	B2	1540	0.9983	50.889	121.591	489.8756
Degree 3	VD16	Total	B2	1540	0.9983	50.889	38.711	156.3199
Degree 4	VD16	Entrena	B2	8855	0.9991	184.342	2.160	9.6222
Degree 4	VD16	Prueba	B2	8855	0.9991	184.342	292.989	1540.2498
Degree 4	VD16	Total	B2	8855	0.9991	184.342	89.423	468.8881
Degree 2	VD17	Entrena	B2	190	0.9890	22.700	8.522	45.8959
Degree 2	VD17	Prueba	B2	190	0.9890	22.700	30.220	138.905
Degree 2	VD17	Total	B2	190	0.9890	22.700	15.032	73.8034
Degree 3	VD17	Entrena	B2	1330	0.9990	51.919	2.224	13.4714
Degree 3	VD17	Prueba	B2	1330	0.9990	51.919	48.668	274.2128
Degree 3	VD17	Total	B2	1330	0.9990	51.919	16.160	91.707
Degree 4	VD17	Entrena	B2	7315	0.9993	164.427	1.910	11.3974
Degree 4	VD17	Prueba	B2	7315	0.9993	164.427	74.101	506.5881
Degree 4	VD17	Total	B2	7315	0.9993	164.427	23.571	159.9797
Degree 2	VD18	Entrena	D3	351	0.9918	23.475	8.045	49.3761
Degree 2	VD18	Prueba	D3	351	0.9918	23.475	592.542	3038.5052
Degree 2	VD18	Total	D3	351	0.9918	23.475	183.424	946.2664
Degree 3	VD18	Entrena	D3	3276	0.9991	78.154	1.368	15.8268
Degree 3	VD18	Prueba	D3	3276	0.9991	78.154	114.653	322.0669
Degree 3	VD18	Total	D3	3276	0.9991	78.154	35.359	107.7143
Degree 4	VD18	Entrena	D3	23751	0.9994	487.838	1.753	12.9194
Degree 4	VD18	Prueba	D3	23751	0.9994	487.838	#####	492670617
Degree 4	VD18	Total	D3	23751	0.9994	487.838	#####	147826165
Degree 2	VD19	Entrena	B2	276	0.9836	41.634	7.308	50.4715
Degree 2	VD19	Prueba	B2	276	0.9836	41.634	36.664	216.3788
Degree 2	VD19	Total	B2	276	0.9836	41.634	16.116	100.2521
Degree 3	VD19	Entrena	B2	2300	0.9980	82.872	2.110	16.4802
Degree 3	VD19	Prueba	B2	2300	0.9980	82.872	117.341	1250.8233

Anexo 6. Resultados de la regresión polinomial (Continuación)

GRADO	NOMBRE	TIPO	YACI-MIENTO	NUMERO DE COEFICIENTES	r ²	TIEMPO (s)	ERROR RELATIVO PROMEDIO (%)	ERROR ABSOLUTO PROMEDIO (bpd)
Degree 3	VD19	Total	B2	2300	0.9980	82.872	36.685	386.8457
Degree 4	VD19	Entrena	B2	14950	0.9983	304.192	1.938	15.3489
Degree 4	VD19	Prueba	B2	14950	0.9983	304.192	170.540	2042.1135
Degree 4	VD19	Total	B2	14950	0.9983	304.192	52.527	623.481
Degree 2	VD20	Entrena	D3	325	0.9902	28.916	5.060	8.8405
Degree 2	VD20	Prueba	D3	325	0.9902	28.916	26.525	28.304
Degree 2	VD20	Total	D3	325	0.9902	28.916	11.501	14.6805
Degree 3	VD20	Entrena	D3	2925	0.9990	84.400	2.113	3.0628
Degree 3	VD20	Prueba	D3	2925	0.9990	84.400	103.936	119.6747
Degree 3	VD20	Total	D3	2925	0.9990	84.400	32.665	38.0523
Degree 4	VD20	Entrena	D3	20475	0.9997	429.151	1.071	1.6384
Degree 4	VD20	Prueba	D3	20475	0.9997	429.151	213.379	280.5088
Degree 4	VD20	Total	D3	20475	0.9997	429.151	64.775	85.3137
Degree 2	VD21	Entrena	B2	231	0.9857	25.942	7.749	49.5764
Degree 2	VD21	Prueba	B2	231	0.9857	25.942	18.886	134.4989
Degree 2	VD21	Total	B2	231	0.9857	25.942	11.091	75.0574
Degree 3	VD21	Entrena	B2	1771	0.9985	64.328	2.177	15.8148
Degree 3	VD21	Prueba	B2	1771	0.9985	64.328	59.920	420.8158
Degree 3	VD21	Total	B2	1771	0.9985	64.328	19.503	137.3357
Degree 4	VD21	Entrena	B2	10626	0.9991	220.147	1.541	12.2389
Degree 4	VD21	Prueba	B2	10626	0.9991	220.147	57.050	524.8346
Degree 4	VD21	Total	B2	10626	0.9991	220.147	18.197	166.0436
Degree 2	VD22	Entrena	E1	171	0.9983	21.064	1.554	7.7573
Degree 2	VD22	Prueba	E1	171	0.9983	21.064	12.053	64.8203
Degree 2	VD22	Total	E1	171	0.9983	21.064	4.704	24.8791
Degree 3	VD22	Entrena	E1	1140	0.9998	48.152	0.518	2.8212
Degree 3	VD22	Prueba	E1	1140	0.9998	48.152	24.474	180.3369
Degree 3	VD22	Total	E1	1140	0.9998	48.152	7.706	56.0849
Degree 4	VD22	Entrena	E1	5985	0.9999	138.072	0.330	1.7427
Degree 4	VD22	Prueba	E1	5985	0.9999	138.072	129.218	686.8422
Degree 4	VD22	Total	E1	5985	0.9999	138.072	39.003	207.3073
Degree 2	VD23	Entrena	D3	253	0.9983	21.043	1.839	1.8463
Degree 2	VD23	Prueba	D3	253	0.9983	21.043	22.972	114.3421
Degree 2	VD23	Total	D3	253	0.9983	21.043	8.180	35.6008

Anexo 6. Resultados de la regresión polinomial (Continuación)

GRADO	NOMBRE	TIPO	YACI-MIENTO	NUMERO DE COEFICIENTES	r ²	TIEMPO (s)	ERROR RELATIVO PROMEDIO (%)	ERROR ABSOLUTO PROMEDIO (bpd)
Degree 3	VD23	Entrena	D3	2024	1.0000	58.733	0.332	0.3553
Degree 3	VD23	Prueba	D3	2024	1.0000	58.733	5892.524	42854.567
Degree 3	VD23	Total	D3	2024	1.0000	58.733	1768.288	12858.7908
Degree 4	VD23	Entrena	D3	12650	1.0000	239.936	0.172	0.1906
Degree 4	VD23	Prueba	D3	12650	1.0000	239.936	55.493	193.8965
Degree 4	VD23	Total	D3	12650	1.0000	239.936	16.771	58.3122
Degree 2	VD24	Entrena	D3	253	0.9862	20.829	11.127	42.8189
Degree 2	VD24	Prueba	D3	253	0.9862	20.829	27.009	193.3847
Degree 2	VD24	Total	D3	253	0.9862	20.829	15.892	87.9963
Degree 3	VD24	Entrena	D3	2024	0.9944	57.799	14.458	32.5233
Degree 3	VD24	Prueba	D3	2024	0.9944	57.799	140.430	862.226
Degree 3	VD24	Total	D3	2024	0.9944	57.799	52.256	281.4762
Degree 4	VD24	Entrena	D3	12650	0.9991	233.470	2.768	11.0785
Degree 4	VD24	Prueba	D3	12650	0.9991	233.470	73.952	534.2806
Degree 4	VD24	Total	D3	12650	0.9991	233.470	24.127	168.0657
Degree 2	VD25	Entrena	E1	231	0.9984	11.817	2.566	5.5834
Degree 2	VD25	Prueba	E1	231	0.9984	11.817	92.209	44.7333
Degree 2	VD25	Total	E1	231	0.9984	11.817	29.474	17.3347
Degree 3	VD25	Entrena	E1	1771	0.9999	33.637	0.452	1.082
Degree 3	VD25	Prueba	E1	1771	0.9999	33.637	1040.230	445.6115
Degree 3	VD25	Total	E1	1771	0.9999	33.637	312.553	134.5126
Degree 4	VD25	Entrena	E1	10626	1.0000	145.081	0.282	0.6116
Degree 4	VD25	Prueba	E1	10626	1.0000	145.081	346.355	485.7657
Degree 4	VD25	Total	E1	10626	1.0000	145.081	104.159	146.2361
Degree 2	VD26	Entrena	E1	210	0.9905	10.611	4.645	8.9314
Degree 2	VD26	Prueba	E1	210	0.9905	10.611	40.041	110.4515
Degree 2	VD26	Total	E1	210	0.9905	10.611	15.262	39.3819
Degree 3	VD26	Entrena	E1	1540	0.9997	30.058	0.730	1.6847
Degree 3	VD26	Prueba	E1	1540	0.9997	30.058	1619.089	6123.6978
Degree 3	VD26	Total	E1	1540	0.9997	30.058	486.150	1837.9554
Degree 4	VD26	Entrena	E1	8855	0.9999	120.998	0.502	1.1276
Degree 4	VD26	Prueba	E1	8855	0.9999	120.998	22033.086	53756.4549
Degree 4	VD26	Total	E1	8855	0.9999	120.998	6609.078	16124.7995
Degree 2	VD27	Entrena	D3	253	0.9996	15.100	1.919	2.3731

Anexo 6. Resultados de la regresión polinomial (Continuación)

GRADO	NOMBRE	TIPO	YACI-MIENTO	NUMERO DE COEFICIENTES	r ²	TIEMPO (s)	ERROR RELATIVO PROMEDIO (%)	ERROR ABSOLUTO PROMEDIO (bpd)
Degree 2	VD27	Prueba	D3	253	0.9996	15.100	15.991	81.3129
Degree 2	VD27	Total	D3	253	0.9996	15.100	6.148	26.0931
Degree 3	VD27	Entrena	D3	2024	1.0000	46.540	0.062	0.108
Degree 3	VD27	Prueba	D3	2024	1.0000	46.540	16.757	97.2689
Degree 3	VD27	Total	D3	2024	1.0000	46.540	5.079	29.3031
Degree 4	VD27	Entrena	D3	12650	1.0000	200.398	0.087	0.1451
Degree 4	VD27	Prueba	D3	12650	1.0000	200.398	775.937	4810.0499
Degree 4	VD27	Total	D3	12650	1.0000	200.398	233.216	1445.4352
Degree 2	VD28	Entrena	E1	253	0.9967	16.106	2.786	4.1529
Degree 2	VD28	Prueba	E1	253	0.9967	16.106	17.045	27.5096
Degree 2	VD28	Total	E1	253	0.9967	16.106	7.049	11.1362
Degree 3	VD28	Entrena	E1	2024	1.0000	49.086	0.022	0.0356
Degree 3	VD28	Prueba	E1	2024	1.0000	49.086	148.024	117.0548
Degree 3	VD28	Total	E1	2024	1.0000	49.086	44.273	35.0226
Degree 4	VD28	Entrena	E1	12650	1.0000	218.975	0.007	0.0108
Degree 4	VD28	Prueba	E1	12650	1.0000	218.975	53.214	106.7455
Degree 4	VD28	Total	E1	12650	1.0000	218.975	15.915	31.9229
Degree 2	VD29	Entrena	D3	210	0.9981	16.033	2.486	15.7862
Degree 2	VD29	Prueba	D3	210	0.9981	16.033	22.301	159.6031
Degree 2	VD29	Total	D3	210	0.9981	16.033	8.429	58.9235
Degree 3	VD29	Entrena	D3	1540	0.9999	40.917	0.667	4.5312
Degree 3	VD29	Prueba	D3	1540	0.9999	40.917	24.686	173.5728
Degree 3	VD29	Total	D3	1540	0.9999	40.917	7.871	55.2346
Degree 4	VD29	Entrena	D3	8855	1.0000	152.729	0.378	2.7264
Degree 4	VD29	Prueba	D3	8855	1.0000	152.729	34.758	274.6922
Degree 4	VD29	Total	D3	8855	1.0000	152.729	10.690	84.3016
Degree 2	VD30	Entrena	D3	231	0.9978	17.716	3.048	16.6249
Degree 2	VD30	Prueba	D3	231	0.9978	17.716	36.954	141.1296
Degree 2	VD30	Total	D3	231	0.9978	17.716	13.220	53.9763
Degree 3	VD30	Entrena	D3	1771	1.0000	53.607	0.315	2.0314
Degree 3	VD30	Prueba	D3	1771	1.0000	53.607	88.273	275.1113
Degree 3	VD30	Total	D3	1771	1.0000	53.607	26.703	83.9554
Degree 4	VD30	Entrena	D3	10626	1.0000	206.519	0.333	1.6586
Degree 4	VD30	Prueba	D3	10626	1.0000	206.519	109.113	794.1633

Anexo 6. Resultados de la regresión polinomial (Continuación)

GRADO	NOMBRE	TIPO	YACI-MIENTO	NUMERO DE COEFICIENTES	r ²	TIEMPO (s)	ERROR RELATIVO PROMEDIO (%)	ERROR ABSOLUTO PROMEDIO (bpd)
Degree 4	VD30	Total	D3	10626	1.0000	206.519	32.967	239.41
Degree 2	VD31	Entrena	E1	210	0.9992	14.848	2.213	6.0696
Degree 2	VD31	Prueba	E1	210	0.9992	14.848	24.469	46.2864
Degree 2	VD31	Total	E1	210	0.9992	14.848	8.880	18.1175
Degree 3	VD31	Entrena	E1	1540	1.0000	42.200	0.223	0.6244
Degree 3	VD31	Prueba	E1	1540	1.0000	42.200	32.411	55.5581
Degree 3	VD31	Total	E1	1540	1.0000	42.200	9.866	17.0811
Degree 4	VD31	Entrena	E1	8855	1.0000	157.980	0.183	0.4382
Degree 4	VD31	Prueba	E1	8855	1.0000	157.980	76.717	134.6067
Degree 4	VD31	Total	E1	8855	1.0000	157.980	23.111	40.6315
Degree 2	VD32	Entrena	E1	190	0.9987	12.516	1.317	3.3187
Degree 2	VD32	Prueba	E1	190	0.9987	12.516	14.465	33.976
Degree 2	VD32	Total	E1	190	0.9987	12.516	5.254	12.4975
Degree 3	VD32	Entrena	E1	1330	1.0000	33.490	0.069	0.1846
Degree 3	VD32	Prueba	E1	1330	1.0000	33.490	295.748	767.257
Degree 3	VD32	Total	E1	1330	1.0000	33.490	88.595	229.847
Degree 4	VD32	Entrena	E1	7315	1.0000	134.708	0.033	0.07
Degree 4	VD32	Prueba	E1	7315	1.0000	134.708	82.930	211.9154
Degree 4	VD32	Total	E1	7315	1.0000	134.708	24.853	63.4968
Degree 2	VD33	Entrena	D3	210	0.9996	15.113	2.654	6.4978
Degree 2	VD33	Prueba	D3	210	0.9996	15.113	28.412	85.6619
Degree 2	VD33	Total	D3	210	0.9996	15.113	10.382	30.247
Degree 3	VD33	Entrena	D3	1540	1.0000	39.549	0.555	1.2932
Degree 3	VD33	Prueba	D3	1540	1.0000	39.549	53.647	166.9663
Degree 3	VD33	Total	D3	1540	1.0000	39.549	16.483	50.9951
Degree 4	VD33	Entrena	D3	8855	1.0000	156.277	0.738	1.4209
Degree 4	VD33	Prueba	D3	8855	1.0000	156.277	167.884	443.343
Degree 4	VD33	Total	D3	8855	1.0000	156.277	50.882	133.9975
Degree 2	VD34	Entrena	C2	253	0.9879	16.986	3.394	25.5955
Degree 2	VD34	Prueba	C2	253	0.9879	16.986	24.539	113.6464
Degree 2	VD34	Total	C2	253	0.9879	16.986	9.735	51.9969
Degree 3	VD34	Entrena	C2	2024	0.9996	50.726	0.764	5.6772
Degree 3	VD34	Prueba	C2	2024	0.9996	50.726	73.541	337.657
Degree 3	VD34	Total	C2	2024	0.9996	50.726	22.586	105.2187

Anexo 6. Resultados de la regresión polinomial (Continuación)

GRADO	NOMBRE	TIPO	YACI-MIENTO	NUMERO DE COEFICIENTES	r ²	TIEMPO (s)	ERROR RELATIVO PROMEDIO (%)	ERROR ABSOLUTO PROMEDIO (bpd)
Degree 4	VD34	Entrena	C2	12650	0.9996	228.851	0.923	5.9079
Degree 4	VD34	Prueba	C2	12650	0.9996	228.851	151.043	857.3915
Degree 4	VD34	Total	C2	12650	0.9996	228.851	45.935	261.2186
Degree 2	VD35	Entrena	C2	210	0.9906	14.461	6.392	19.6002
Degree 2	VD35	Prueba	C2	210	0.9906	14.461	45.793	120.6227
Degree 2	VD35	Total	C2	210	0.9906	14.461	18.227	49.9442
Degree 3	VD35	Entrena	C2	1540	0.9999	42.852	0.608	2.1783
Degree 3	VD35	Prueba	C2	1540	0.9999	42.852	96.458	249.5199
Degree 3	VD35	Total	C2	1540	0.9999	42.852	29.398	76.4719
Degree 4	VD35	Entrena	C2	8855	0.9999	173.189	0.433	1.4907
Degree 4	VD35	Prueba	C2	8855	0.9999	173.189	465.884	1289.7642
Degree 4	VD35	Total	C2	8855	0.9999	173.189	140.240	388.4471
Degree 2	VD36	Entrena	C2	231	0.9962	14.539	2.986	13.6142
Degree 2	VD36	Prueba	C2	231	0.9962	14.539	15.028	70.9535
Degree 2	VD36	Total	C2	231	0.9962	14.539	6.594	30.7944
Degree 3	VD36	Entrena	C2	1771	1.0000	45.546	0.296	1.5705
Degree 3	VD36	Prueba	C2	1771	1.0000	45.546	36.484	154.2801
Degree 3	VD36	Total	C2	1771	1.0000	45.546	11.138	47.3258
Degree 4	VD36	Entrena	C2	10626	1.0000	198.413	0.344	1.6853
Degree 4	VD36	Prueba	C2	10626	1.0000	198.413	66.959	273.7118
Degree 4	VD36	Total	C2	10626	1.0000	198.413	20.303	83.1908
Degree 2	VD37	Entrena	C2	171	0.9959	15.686	4.587	1.2803
Degree 2	VD37	Prueba	C2	171	0.9959	15.686	88.820	14.1775
Degree 2	VD37	Total	C2	171	0.9959	15.686	29.852	5.1488
Degree 3	VD37	Entrena	C2	1140	1.0000	37.699	0.057	0.0213
Degree 3	VD37	Prueba	C2	1140	1.0000	37.699	799.989	123.8803
Degree 3	VD37	Total	C2	1140	1.0000	37.699	239.993	37.1722
Degree 4	VD37	Entrena	C2	5985	1.0000	121.327	0.002	0.0006
Degree 4	VD37	Prueba	C2	5985	1.0000	121.327	11473.627	2629.9346
Degree 4	VD37	Total	C2	5985	1.0000	121.327	3441.465	788.8376
Degree 2	VD38	Entrena	C2	190	0.9992	13.912	0.682	1.127
Degree 2	VD38	Prueba	C2	190	0.9992	13.912	0.000	0
Degree 2	VD38	Total	C2	190	0.9992	13.912	0.478	0.7894
Degree 3	VD38	Entrena	C2	1330	1.0000	35.868	0.002	0.0038

Anexo 6. Resultados de la regresión polinomial (Continuación)

GRADO	NOMBRE	TIPO	YACI- MIEN- TO	NUME- RO DE COEFI- CIEN- TES	r ²	TIEMPO (s)	ERROR RELATIVO PROMEDIO (%)	ERROR ABSOLUTO PROMEDIO (bpd)
Degree 3	VD38	Prueba	C2	1330	1.0000	35.868	0.000	0
Degree 3	VD38	Total	C2	1330	1.0000	35.868	0.002	0.0026
Degree 4	VD38	Entrena	C2	7315	1.0000	140.839	0.001	0.002
Degree 4	VD38	Prueba	C2	7315	1.0000	140.839	0.000	0
Degree 4	VD38	Total	C2	7315	1.0000	140.839	0.001	0.0014
Degree 2	VD39	Entrena	C2	190	0.9993	13.728	1.093	1.0743
Degree 2	VD39	Prueba	C2	190	0.9993	13.728	59.604	10.6364
Degree 2	VD39	Total	C2	190	0.9993	13.728	18.662	3.9455
Degree 3	VD39	Entrena	C2	1330	1.0000	38.665	0.043	0.0421
Degree 3	VD39	Prueba	C2	1330	1.0000	38.665	754.074	106.7063
Degree 3	VD39	Total	C2	1330	1.0000	38.665	226.457	32.0703
Degree 4	VD39	Entrena	C2	7315	1.0000	140.491	0.005	0.0043
Degree 4	VD39	Prueba	C2	7315	1.0000	140.491	112.998	21.8268
Degree 4	VD39	Total	C2	7315	1.0000	140.491	33.933	6.557
Degree 2	VD40	Entrena	C2	231	0.9992	11.900	2.583	7.0209
Degree 2	VD40	Prueba	C2	231	0.9992	11.900	7.780	29.8716
Degree 2	VD40	Total	C2	231	0.9992	11.900	4.143	13.8823
Degree 3	VD40	Entrena	C2	1771	0.9999	35.296	0.974	2.0789
Degree 3	VD40	Prueba	C2	1771	0.9999	35.296	109.217	436.6233
Degree 3	VD40	Total	C2	1771	0.9999	35.296	33.476	132.56
Degree 4	VD40	Entrena	C2	10626	1.0000	154.244	0.719	1.5101
Degree 4	VD40	Prueba	C2	10626	1.0000	154.244	31.868	228.4562
Degree 4	VD40	Total	C2	10626	1.0000	154.244	10.072	69.6554

Anexo 7: Resultados de redes neuronales (Set de datos normalizados caracterizados, función de activación MAE)

Anexo 7. Resultados de redes neuronales

Nombre del pozo	Capas	Numero de neuronas	Error absoluto promedio (bpd / día)	Error relativo promedio (%)	Tiempo (s)	Ambiente	Yacimiento
VD01	1	10	145.45	33.4870	2325.36	DELTAIC	B2
VD02	1	13	109.73	34.9099	2287.19	FLUVIAL	D3
VD03	3	50	65.20	10.8503	2315.29	DELTAIC	B2
VD04	3	50	118.38	19.8059	2324.43	FLUVIAL	D3
VD05	3	41	55.21	12.7655	2289.91	DELTAIC	B2
VD06	1	8	73.61	19.6517	2269.85	FLUVIAL	D3
VD07	3	12	107.16	15.3264	2266.36	DELTAIC	B2
VD08	3	39	58.35	33.0327	2304.45	FLUVIAL	D3
VD09	3	51	102.40	20.1316	2305.86	DELTAIC	B2
VD10	3	50	28.49	9.8766	2349.50	DELTAIC	C2
VD11	2	200	52.35	23.1037	2333.33	FLUVIAL	D3
VD12	2	41	79.91	18.3216	2279.92	DELTAIC	B2
VD13	3	10	130.10	68.3126	2278.65	FLUVIAL	D3
VD14	3	50	68.92	20.5020	2268.99	DELTAIC	B2
VD15	3	12	84.51	38.7405	2261.92	DELTAIC	B2
VD16	2	200	104.33	55.3419	2284.96	DELTAIC	B2
VD17	3	8	220.53	42.7659	2282.66	DELTAIC	B2
VD18	1	16	766.11	64.8510	2517.60	FLUVIAL	D3
VD19	1	11	118.63	22.4972	2440.05	DELTAIC	B2
VD20	3	50	27.72	20.4787	2451.53	FLUVIAL	D3
VD21	1	48	146.72	24.1030	2536.52	DELTAIC	B2
VD22	2	11	42.11	7.8880	2422.14	FLUVIAL	E1
VD23	3	37	185.93	101.4082	2379.74	FLUVIAL	D3
VD24	3	50	86.19	26.0860	1945.31	FLUVIAL	D3
VD25	2	200	20.07	12.3082	939.68	FLUVIAL	E1
VD26	2	200	86.42	33.9213	951.93	FLUVIAL	E1
VD27	2	20	56.71	23.1459	957.24	FLUVIAL	D3
VD28	3	50	17.60	12.2084	963.08	FLUVIAL	E1
VD29	3	200	39.60	6.2680	957.30	FLUVIAL	D3
VD30	2	200	22.52	3.7786	966.31	FLUVIAL	D3
VD31	1	50	23.71	11.4498	970.09	FLUVIAL	E1

Anexo 7. Resultados de redes neuronales (Continuación)

Nombre del pozo	Capas	Numero de neuronas	Error absoluto promedio (bpd / día)	Error relativo promedio (%)	Tiempo (s)	Ambiente	Yacimiento
VD32	2	25	18.93	7.1587	954.70	FLUVIAL	E1
VD33	1	50	73.29	32.2098	986.46	FLUVIAL	D3
VD34	3	14	109.92	18.8520	1005.49	DELTAIC	C2
VD35	1	9	77.98	33.5995	1009.24	DELTAIC	C2
VD36	3	35	68.47	13.1056	984.88	DELTAIC	C2
VD37	2	200	4.40	11.7177	974.76	DELTAIC	C2
VD39	3	31	5.16	14.2267	971.24	DELTAIC	C2
VD40	3	13	28.80	11.8560	976.75	DELTAIC	C2

Anexo 8: Comparación de resultados de los modelos en cada pozo

Anexo 8. Resultados de todos los modelos de predicción

POZO	LINEAL		REDES NEURONALES		GRADO 2		GRADO 3		GRADO 4	
	RELATIVO %	ABSOLUTO (stb / day)	RELATIVO %2	ABSOLUTO (stb / day)3	RELATIVO %4	ABSOLUTO (stb / day)5	RELATIVO %6	ABSOLUTO (stb / day)7	RELATIVO %8	ABSOLUTO (stb / day)9
VD01	25.74%	148	33.49%	145	33.63%	105	39.31%	147	68.39%	373
VD02	31.24%	105	34.91%	110	12.27%	46	50.09%	90	317.85%	231
VD03	21.99%	131	10.85%	65	11.30%	66	16.97%	84	27.36%	124
VD04	22.19%	137	19.81%	118	12.32%	80	29.95%	154	35.53%	256
VD05	24.80%	118	12.77%	55	13.27%	54	17.21%	88	17.48%	83
VD06	28.36%	104	19.65%	74	12.64%	60	20.75%	86	79.65%	200
VD07	22.90%	123	15.33%	107	10.49%	83	30.56%	246	26.05%	266
VD08	22.74%	52	33.03%	58	12.80%	28	23.47%	66	111.69%	177
VD09	22.90%	114	20.13%	102	25.27%	82	64.05%	175	53.88%	131
VD10	9.03%	26	9.88%	28	3.66%	10	54.45%	113	69.92%	176
VD11	20.18%	58	23.10%	52	13.08%	38	30.72%	70	38.84%	125
VD12	16.30%	83	18.32%	80	7.24%	42	22.31%	129	41.68%	252
VD13	147.00%	132	68.31%	130	14.14%	72	47.71%	205	30.63%	188
VD14	17.39%	76	20.50%	69	11.93%	52	25.05%	111	129.04%	551
VD15	25.42%	60	38.74%	85	17.47%	39	25.86%	57	64.15%	95
VD16	28.21%	100	55.34%	104	99.46%	162	38.71%	156	89.42%	469
VD17	21.71%	121	42.77%	221	15.03%	74	16.16%	92	23.57%	160
VD18	41.68%	276	64.85%	766	183.42%	946	35.36%	108	2.96E+05	1.48E+08
VD19	18.54%	125	22.50%	119	16.12%	100	36.69%	387	52.53%	623
VD20	15.52%	25	20.48%	28	11.50%	15	32.67%	38	64.77%	85
VD21	17.53%	118	24.10%	147	11.09%	75	19.50%	137	18.20%	166
VD22	4.34%	27	7.89%	42	4.70%	25	7.71%	56	39.00%	207
VD23	12.53%	14	101.41%	186	8.18%	36	1768.29%	12859	16.77%	58
VD24	23.79%	113	26.09%	86	15.89%	88	52.26%	281	24.13%	168
VD25	101.88%	37	12.31%	20	29.47%	17	312.55%	135	104.16%	146
VD26	27.40%	30	33.92%	86	15.26%	39	486.15%	1838	6609.08%	16125
VD27	28.32%	49	23.15%	57	6.15%	26	5.08%	29	233.22%	1445
VD28	13.57%	18	12.21%	18	7.05%	11	44.27%	35	15.91%	32
VD29	18.16%	112	6.27%	40	8.43%	59	7.87%	55	10.69%	84
VD30	14.12%	73	3.78%	23	13.22%	54	26.70%	84	32.97%	239
VD31	22.45%	54	11.45%	24	8.88%	18	9.87%	17	23.11%	41

Anexo 8. Resultados de todos los modelos de predicción (Continuación)

POZO	LINEAL		REDES NEURONALES		GRADO 2		GRADO 3		GRADO 4	
	RELATIVO %	ABSOLUTO (stb / day)	RELATIVO %2	ABSOLUTO (stb / day)3	RELATIVO %4	ABSOLUTO (stb / day)5	RELATIVO %6	ABSOLUTO (stb / day)7	RELATIVO %8	ABSOLUTO (stb / day)9
VD32	7.73%	19	7.16%	19	5.25%	12	88.60%	230	24.85%	63
VD33	47.46%	82	32.21%	73	10.38%	30	16.48%	51	50.88%	134
VD34	13.81%	86	18.85%	110	9.73%	52	22.59%	105	45.94%	261
VD35	16.00%	48	33.60%	78	18.23%	50	29.40%	76	140.24%	388
VD36	13.80%	58	13.11%	68	6.59%	31	11.14%	47	20.30%	83
VD37	129.34%	21	11.72%	4	29.85%	5	239.99%	37	3441.46%	789
VD38	1.57%	3	Error	Error	0.48%	1	0.00%	0	0.00%	0
VD39	29.17%	9	14.23%	5	18.66%	4	226.46%	32	33.93%	7
VD40	10.11%	39	11.86%	29	4.14%	14	33.48%	133	10.07%	70

Anexo 9: Resultado del análisis de bombas y su significancia estadística por pozo y modelo de bomba

Anexo 9. Resultados del modelo de análisis de bombas BCP

Wellname	Pump_Model	Pump No.	P values	Activ e Pump Days	Pump Displacement	% Real Production	Pump Displacement real
VD01	750TP1200	Pump 1	0	861	8,9975	0,6736	6,060541191
VD01	1000TP860	Pump 2	0,0004	101	12,54	0,4979	6,243347128
VD01	200-1350	Pump 3	0	620	12,6	0,2873	3,620156888
VD01	200-1350	Pump 4	0	132	12,6	0,2342	2,95085119
VD01	1000TP860	Pump 5	0,0004	1042	12,54	0,2282	2,86158488
VD01	300TP1800	Pump 6	0	566	3,61	0,6139	2,216224673
VD01	430TP2000	Pump 7	0	231	5,1575	0,3737	1,927520813
VD02	430TP2000	Pump 1	0	424	5,1575	0,7525	3,881036366
VD02	175-1200	Pump 2	0,9363	549	0		
VD02	1000TP860	Pump 3	0	402	12,54	0,4167	5,225779855
VD02	175-1200	Pump 4	0,9363	136	0		
VD02	1000TP860	Pump 5	0	68	12,54	0,2339	2,933307676
VD02	1000TP860	Pump 6	0	444	12,54	0,2722	3,41336456
VD02	-	Pump 7	0,0002	918	0		
VD02	1000TP860	Pump 8	0	380	12,54	0,0499	0,625756536
VD02	1000TP860	Pump 9	0	42	12,54	0,0348	0,435880617
VD02	200-1350	Pump 10	0	4	12,6	0,0235	0,295679287
VD02	200-1350	Pump 11	0	8	12,6	0,0241	0,30420376
VD03	430TP2000	Pump 1	0	362	5,1575	0,7331	3,781179674
VD03	1000TP860	Pump 2	0	410	12,54	0,5386	6,753581248
VD03	1000TP860	Pump 3	0	539	12,54	0,3307	4,146390603
VD03	1000TP860	Pump 4	0	607	12,54	0,2239	2,807792198
VD03	1000TP860	Pump 5	0	496	12,54	0,1832	2,297862493

Anexo 9. Resultados del modelo de análisis de bombas BCP (Continuación)

Wellname	Pump_Model	Pump No.	P values	Active Pump Days	Pump Displacement	% Real Production	Pump Displacement real
VD03	1000TP860	Pump 6	0	928	12,54	0,2002	2,51011662
VD03	1000TP860	Pump 7	0	14	12,54	0,2985	3,742627259
VD04	-	Pump 1	0	554	0		
VD04	1000TP860	Pump 2	0	450	12,54	0,5869	7,360345774
VD04	1000TP860	Pump 3	0	407	12,54	0,4719	5,918215021
VD04	165-1350	Pump 4	0,3696	16	10,3	0,3993	4,112625654
VD04	175-1200	Pump 5	0	169	0		
VD04	1000TP860	Pump 6	0	116	12,54	0,1955	2,451152658
VD04	1000TP860	Pump 7	0	153	12,54	0,1868	2,342664945
VD04	750TP1200	Pump 8	0	659	8,9975	0,3604	3,243095712
VD04	580TP1600	Pump 9	0	912	6,8875	0,3296	2,270462175
VD05	430TP2000	Pump 1	0,0001	198	5,1575	0,7759	4,001506314
VD05	175-1200	Pump 2	0	335	0		
VD05	200-1350	Pump 3	0	333	12,6	0,4012	5,055111401
VD05	1000TP860	Pump 4	0	679	12,54	0,3299	4,137373101
VD05	1000TP860	Pump 5	0	151	12,54	0,2209	2,770411272
VD05	1000TP860	Pump 6	0	567	12,54	0,2019	2,531567599
VD05	20.40-2100	Pump 7	0	687	4,2	0,2055	0,863186416
VD05	20.40-2100	Pump 8	0	286	4,2	0,4263	1,790463112
VD05	580TP1600	Pump 9	0	234	6,8875	0,2551	1,756835949
VD06	430TP2000	Pump 1	0	417	5,1575	0,6974	3,596671452
VD06	1000TP860	Pump 2	0	843	12,54	0,5150	6,458625249
VD06	1000TP860	Pump 3	0	252	12,54	0,2134	2,675717991
VD06	1000TP860	Pump 4	0	234	12,54	0,1517	1,901899015
VD06	1000TP860	Pump 5	0	663	12,54	0,2049	2,569981644
VD06	1000TP860	Pump 6	0	1008	12,54	0,1619	2,029930798
VD07	750TP1200	Pump 1	0	1221	8,9975	0,5982	5,382578631
VD07	1000TP860	Pump 2	0	473	12,54	0,2545	3,191765056
VD07	1000TP860	Pump 3	0	883	12,54	0,2517	3,156209541
VD07	1000TP860	Pump 4	0	629	12,54	0,2210	2,771227691
VD07	580TP1600	Pump 5	0	324	6,8875	0,3469	2,389589483

Anexo 9. Resultados del modelo de análisis de bombas BCP (Continuación)

Wellname	Pump_Model	Pump No.	P values	Activ e Pump Days	Pump Displacement	% Real Production	Pump Displacement real
VD08	430TP2000	Pump 1	0	380	5,1575	0,7353	3,792547611
VD08	1000TP860	Pump 2	0,8366	418	12,54	0,6798	8,524414243
VD08	750TP1200	Pump 3	0	1615	8,9975	0,4046	3,640065684
VD08	-	Pump 4	0,0204	65	0		
VD08	20.40-2100	Pump 5	0,4839	297	4,2	0,2125	0,892402238
VD08	20.40-2100	Pump 6	0,4839	320	4,2	0,0843	0,35404148
VD08	175-1200	Pump 7	0,0094	0	0		
VD09	430TP2000	Pump 1	0,3249	219	5,1575	0,7712	3,97726474
VD09	750TP1200	Pump 2	0	871	8,9975	0,6000	5,398364631
VD09	1000TP860	Pump 3	0,0015	543	12,54	0,2934	3,67907554
VD09	1000TP860	Pump 4	0,0015	116	12,54	0,1719	2,155674847
VD09	1000TP860	Pump 5	0,0015	793	12,54	0,1882	2,359612241
VD09	1000TP860	Pump 6	0,0015	997	12,54	0,2668	3,345920671
VD10	430TP2000	Pump 1	0	339	5,1575	0,5914	3,05006431
VD10	24.40-1500	Pump 2	0	1175	3	0,7162	2,148641488
VD10	1000TP860	Pump 3	0	159	12,54	0,0723	0,907092824
VD10	20.40-2100	Pump 4	0	149	4,2	0,1735	0,728601619
VD11	430TP2000	Pump 1	0	933	5,1575	0,5710	2,945042425
VD11	98-1350	Pump 2	0,0225	43	6,15	0,7411	4,55746407
VD11	750TP1200	Pump 3	0	678	8,9975	0,3608	3,246437599
VD11	580TP1600	Pump 4	0	1077	6,8875	0,2633	1,813523158
VD11	98-1350	Pump 5	0,0225	107	6,15	0,1052	0,647281116
VD11	750TP1200	Pump 6	0	576	8,9975	0,1591	1,431771963
VD12	430TP2000	Pump 1	0,735	940	5,1575	0,7169	3,69737978
VD12	1000TP860	Pump 2	0	430	12,54	0,4061	5,09244603
VD12	-	Pump 3	0	811	0		
VD12	22.40-2500	Pump 4	0	748	5	0,4054	2,027095885
VD12	1000TP860	Pump 5	0	74	12,54	0,2530	3,172973321
VD12	-	Pump 6	0	405	0		
VD12	580TP1600	Pump 7	0	28	6,8875	0,5374	3,701419751
VD12	175-1200	Pump 8	0	119	0		
VD13	-	Pump 1	0,091	529	0		
VD13	1000TP860	Pump 2	0,0019	373	12,54	0,6061	7,600246281
VD13	1000TP860	Pump 3	0,0019	985	12,54	0,2842	3,56368679
VD13	750TP1200	Pump 4	0	40	8,9975	0,3404	3,062715642
VD13	175-1200	Pump 5	0,0978	515	0		

Anexo 9. Resultados del modelo de análisis de bombas BCP (Continuación)

Wellname	Pump_Model	Pump No.	P values	Activ e Pump Days	Pump Displacement	% Real Production	Pump Displacement real
VD13	1000TP860	Pump 6	0,0019	248	12,54	0,0874	1,095966594
VD13	165-1350	Pump 7	0	412	10,3	0,0823	0,847708283
VD13	1000TP860	Pump 8	0,0019	124	12,54	0,0318	0,398668413
VD13	1000TP860	Pump 9	0,0019	72	12,54	0,0407	0,510491829
VD13	1000TP860	Pump 10	0,0019	14	12,54	0,0643	0,806675458
VD14	430TP2000	Pump 1	0,0585	585	5,1575	0,6790	3,50199516
VD14	-	Pump 2	0,0002	458	0		
VD14	1000TP860	Pump 3	0	485	12,54	0,2542	3,187084785
VD14	1000TP860	Pump 4	0	322	12,54	0,1468	1,840479564
VD14	165-1350	Pump 5	0	228	10,3	0,0867	0,892783649
VD14	1000TP860	Pump 6	0	937	12,54	0,2039	2,556505506
VD14	750TP1200	Pump 7	0,0001	555	8,9975	0,4789	4,308857951
VD15	430TP2000	Pump 1	0	438	5,1575	0,6804	3,509098408
VD15	750TP1200	Pump 2	0,8571	1704	8,9975	0,4615	4,152762694
VD15	300TP1800	Pump 3	0	476	3,61	0,2574	0,929320985
VD15	300TP1800	Pump 4	0	451	3,61	0,1629	0,588187078
VD15	300TP1800	Pump 5	0	388	3,61	0,5165	1,864468853
VD16	-	Pump 1	0	464	0		
VD16	1000TP860	Pump 2	0	852	12,54	0,4413	5,534170442
VD16	1000TP860	Pump 3	0	1601	12,54	0,2954	3,703750732
VD16	200TP1800	Pump 4	0	600	2,3375	0,6446	1,506747685
VD16	200TP1800	Pump 5	0	10	2,3375	0,3856	0,901352674
VD17	750TP1200	Pump 1	0,0001	649	8,9975	0,7147	6,430193732
VD17	1000TP860	Pump 2	0,0001	733	12,54	0,4044	5,071771973
VD17	1000TP860	Pump 3	0,0001	361	12,54	0,1753	2,197792067
VD17	1000TP860	Pump 4	0,0001	320	12,54	0,1330	1,667213109
VD17	1000TP860	Pump 5	0,0001	1303	12,54	0,2083	2,61218579
VD17	1000TP860	Pump 6	0,0001	77	12,54	0,2372	2,97412281
VD17	750TP1200	Pump 7	0,0001	86	8,9975	0,2220	1,997587836
VD18	750TP1200	Pump 1	0	161	8,9975	0,6141	5,525720322
VD18	1000TP860	Pump 2	0,0001	524	12,54	0,6927	8,686290446
VD18	-	Pump 3	0	217	0		
VD18	1000TP860	Pump 4	0,0001	326	12,54	0,5345	6,702876696
VD18	1000TP860	Pump 5	0,0001	413	12,54	0,4193	5,257453464
VD18	1000TP860	Pump 6	0,0001	133	12,54	0,4010	5,028346964

Anexo 9. Resultados del modelo de análisis de bombas BCP (Continuación)

Wellname	Pump_Model	Pump No.	P values	Activ e Pump Days	Pump Displacement	% Real Production	Pump Displacement real
VD18	165-1350	Pump 7	0,0018	35	10,3	0,4121	4,244260409
VD18	1000TP860	Pump 8	0,0001	350	12,54	0,3319	4,161797156
VD18	175-1200	Pump 9	0	151	0		
VD18	175-1200	Pump 10	0	196	0		
VD18	130-1200	Pump 11	0	82	0		
VD18	1000TP860	Pump 12	0,0001	253	12,54	0,3753	4,706841389
VD18	NTZ500*120 DT226	Pump 13	0,0026	24	14,21	0,1616	2,296549079
VD18	165-1350	Pump 14	0,0018	192	10,3	0,1576	1,62277079
VD18	-	Pump 15	0	137	0		
VD18	200-1350	Pump 16	0	136	12,6	0,1719	2,165591194
VD18	1000TP860	Pump 17	0,0001	99	12,54	0,1379	1,728730567
VD18	430TP2000	Pump 18	0	98	5,1575	0,3103	1,600195017
VD19	430TP2000	Pump 1	0	149	5,1575	0,7349	3,790350308
VD19	750TP1200	Pump 2	0	910	8,9975	0,6171	5,551928487
VD19	1000TP860	Pump 3	0	421	12,54	0,3811	4,77959185
VD19	200-1350	Pump 4	0	74	12,6	0,2502	3,152812141
VD19	1000TP860	Pump 5	0	839	12,54	0,2407	3,017880894
VD19	22.40-2500	Pump 6	0,1016	863	5	0,4682	2,341235082
VD19	1000TP860	Pump 7	0	115	12,54	0,1659	2,080604472
VD19	-	Pump 8	0	106	0		
VD20	430TP2000	Pump 1	0	540	5,1575	0,7121	3,672473198
VD20	24.40-1200	Pump 2	0	185	2,4	0,4916	1,179810874
VD20	20.40-2100	Pump 3	0,0078	258	4,2	0,2170	0,911606958
VD20	580TP1600	Pump 4	0	313	6,8875	0,1953	1,345236493
VD20	22.40-2500	Pump 5	0	274	5	0,1373	0,686738847
VD20	20.40-2100	Pump 6	0,0078	1126	4,2	0,2516	1,05664664
VD20	430TP2000	Pump 7	0	38	5,1575	0,0919	0,473911369
VD20	750TP1200	Pump 8	0,0001	18	8,9975	0,0518	0,466354098
VD20	430TP2000	Pump 9	0	36	5,1575	0,0930	0,479787817
VD20	-	Pump 10	0,6959	20	0		
VD20	28.40-500	Pump 11	0,0021	65	1,07	0,4937	0,528302297
VD21	750TP1200	Pump 1	0	938	8,9975	0,6576	5,917182773
VD21	200-1350	Pump 2	0	46	12,6	0,3804	4,793398975
VD21	1000TP860	Pump 3	0,5157	128	12,54	0,4127	5,175322271
VD21	1000TP860	Pump 4	0,5157	460	12,54	0,3727	4,673411594
VD21	175-1200	Pump 5	0	457	0		

Anexo 9. Resultados del modelo de análisis de bombas BCP (Continuación)

Wellname	Pump_Model	Pump No.	P values	Activ e Pump Days	Pump Displacement	% Real Production	Pump Displacement real
VD21	175-1200	Pump 6	0	389	0		
VD21	175-1200	Pump 7	0	893	0		
VD21	1000TP860	Pump 8	0,5157	124	12,54	0,1860	2,332542434
VD21	750TP1200	Pump 9	0	91	8,9975	0,2669	2,401213331
VD22	750TP1200	Pump 1	0	1012	8,9975	0,6512	5,859150855
VD22	750TP1200	Pump 2	0	610	8,9975	0,4129	3,714873883
VD23	430TP2000	Pump 1	0	130	5,1575	0,6521	3,363046537
VD23	200-1350	Pump 2	0	727	12,6	0,2566	3,232703384
VD23	24.40-1200	Pump 3	0	630	2,4	0,3105	0,745158756
VD23	200TP1800	Pump 4	0,0032	17	2,3375	0,2788	0,65168134
VD23	-	Pump 5	0,0001	169	0		
VD23	200TP1800	Pump 6	0,0032	83	2,3375	0,1298	0,303348598
VD24	750TP1200	Pump 1	0	386	8,9975	0,6990	6,288812012
VD24	1000TP860	Pump 2	0	519	12,54	0,5372	6,736962041
VD24	1000TP860	Pump 3	0	137	12,54	0,2855	3,579617877
VD24	1000TP860	Pump 4	0	377	12,54	0,3111	3,901646881
VD24	1000TP860	Pump 5	0	428	12,54	0,1708	2,142267798
VD24	175-1200	Pump 6	0,075	94	0		
VD24	22.40-2500	Pump 7	0	387	5	0,4818	2,409163976
VD24	22.40-2500	Pump 8	0	269	5	0,2995	1,497735268
VD24	22.40-2500	Pump 9	0	124	5	0,1823	0,911315407
VD24	22.40-2500	Pump 10	0	128	5	0,0982	0,491181461
VD24	-	Pump 11	0	6	0		
VD24	175-1200	Pump 12	0,075	299	0		
VD24	175-1200	Pump 13	0,075	52	0		
VD25	200-1350	Pump 1	0	24	12,6	0,5181	6,52787063
VD25	165-1350	Pump 2	0	791	10,3	0,3368	3,468995645
VD25	1000TP860	Pump 3	0	804	12,54	0,0595	0,74605351
VD25	1000TP860	Pump 4	0	72	12,54	0,0460	0,576768617
VD25	200-1350	Pump 5	0	79	12,6	0,0355	0,447636163
VD25	-	Pump 6	0,361	3	0		
VD26	200-1350	Pump 1	0,1109	800	12,6	0,4072	5,130563101
VD26	430TP2000	Pump 2	0,1139	190	5,1575	0,1450	0,747905368
VD26	175-1200	Pump 3	0,1797	785	0		
VD27	750TP1200	Pump 1	0	275	8,9975	0,3646	3,280482519
VD27	1000TP860	Pump 2	0	131	12,54	0,1288	1,615102957

Anexo 9. Resultados del modelo de análisis de bombas BCP (Continuación)

Wellname	Pump_Model	Pump No.	P values	Activ e Pump Days	Pump Displacement	% Real Production	Pump Displacement real
VD27	24.40-1500	Pump 3	0	674	3	0,3606	1,081913492
VD27	-	Pump 4	0,0253	391	0		
VD27	580TP1600	Pump 5	0,0327	135	6,8875	0,3470	2,389966722
VD27	750TP1200	Pump 6	0	88	8,9975	0,1752	1,575960043
VD27	1000TP860	Pump 7	0	79	12,54	0,0173	0,21695299
VD28	200-1350	Pump 1	0,0003	186	12,6	0,2856	3,599068561
VD28	750TP1200	Pump 2	0	445	8,9975	0,1427	1,284364684
VD28	175-1200	Pump 3	0	379	0		
VD28	175-1200	Pump 4	0	612	0		
VD28	-	Pump 5	0	114	0		
VD28	400TP1350	Pump 6	0	54	4,8175	0,1419	0,683426424
VD29	200-1350	Pump 1	0,2746	162	12,6	0,5019	6,323909051
VD29	1000TP860	Pump 2	0,2994	253	12,54	0,4663	5,846973434
VD29	-	Pump 3	0,9925	429	0		
VD29	-	Pump 4	0,9925	177	0		
VD29	1000TP860	Pump 5	0,2994	131	12,54	0,1905	2,388353007
VD29	1000TP860	Pump 6	0,2994	89	12,54	0,1679	2,105116137
VD29	-	Pump 7	0,9925	167	0		
VD29	200-1350	Pump 8	0,2746	63	12,6	0,2358	2,970976395
VD29	1000TP860	Pump 9	0,2994	323	12,54	0,3039	3,81042472
VD30	200-1350	Pump 1	0,2894	544	12,6	0,6055	7,628783375
VD30	430TP2000	Pump 2	0,7184	330	5,1575	0,7653	3,94679297
VD30	1000TP860	Pump 3	0,2098	78	12,54	0,3582	4,491294193
VD30	1000TP860	Pump 4	0,2098	137	12,54	0,3360	4,213135221
VD30	165-1350	Pump 5	0,9214	580	10,3	0,4138	4,26192498
VD30	1000TP860	Pump 6	0,2098	97	12,54	0,4729	5,929779656
VD31	200-1350	Pump 1	0	864	12,6	0,4215	5,311486039
VD31	430TP2000	Pump 2	0	647	5,1575	0,2727	1,406684765
VD31	-	Pump 3	0	325	0		
VD32	750TP1200	Pump 1	0,0355	357	8,9975	0,6377	5,737783757
VD32	20.40-2100	Pump 2	0,0366	395	4,2	0,3235	1,358767032
VD32	20.40-2100	Pump 3	0,0366	32	4,2	0,2736	1,149188319
VD33	200-1350	Pump 1	0	192	12,6	0,6539	8,239126323
VD33	750TP1200	Pump 2	0,0014	1046	8,9975	0,3038	2,733755007
VD33	-	Pump 3	0	596	0		
VD34	200-1350	Pump 1	0,0011	1152	12,6	0,5403	6,808240805

Anexo 9. Resultados del modelo de análisis de bombas BCP (Continuación)

Wellname	Pump_Model	Pump No.	P values	Active Pump Days	Pump Displacement	% Real Production	Pump Displacement real
VD34	98-1600	Pump 2	0	84	6,15	0,7381	4,539190086
VD34	750TP1200	Pump 3	0,2944	94	8,9975	0,3564	3,206970341
VD34	1000TP860	Pump 4	0	24	12,54	0,1749	2,193573232
VD34	175-1200	Pump 5	0	73	0		
VD34	200-1350	Pump 6	0,0011	149	12,6	0,4621	5,822122296
VD34	1000TP860	Pump 7	0	252	12,54	0,2911	3,64981165
VD35	200-1350	Pump 1	0,0001	1377	12,6	0,4131	5,205140764
VD35	175-1200	Pump 2	0,341	57	0		
VD35	750TP1200	Pump 3	0,0001	89	8,9975	0,1537	1,38254962
VD35	750TP1200	Pump 4	0,0001	326	8,9975	0,2415	2,173108433
VD36	200-1350	Pump 1	0,0001	72	12,6	0,4888	6,158288105
VD36	1000TP860	Pump 2	0	110	12,54	0,4176	5,236801922
VD36	750TP1200	Pump 3	0	1441	8,9975	0,5103	4,591503895
VD36	1000TP860	Pump 4	0	200	12,54	0,1685	2,112647328
VD36	-	Pump 5	0	0	0		
VD37	24.40-1500	Pump 1	0,0024	1335	3	0,2437	0,731038578
VD38	430TP2000	Pump 1	0,8922	207	5,1575	0,4667	2,407075016
VD38	430TP2000	Pump 2	0,8922	0	5,1575		
VD38	-	Pump 3	0,7631	92	0		
VD39	20.40-2100	Pump 1	0	354	4,2	0,4461	1,87381657
VD39	60TP1300	Pump 2	0	996	0,77	0,4858	0,374052608
VD40	580TP1600	Pump 1	0	12	6,8875	0,7045	4,852353466
VD40	165-1350	Pump 2	0	64	10,3	0,4581	4,718312015
VD40	1000TP860	Pump 3	0	271	12,54	0,5362	6,723817343
VD40	-	Pump 4	0	428	0		
VD40	-	Pump 5	0	0	0		
VD40	-	Pump 6	0	237	0		
VD40	-	Pump 7	0	6	0		

Anexo 10: Manual para el usuario de la herramienta *AJ Analytic tool*



MANUAL DEL USUARIO

DESARROLLADO POR:

Mijares, Adriana
Sánchez, John

AJ Analytic Tool es una herramienta de análisis de datos creada para la industria petrolera, la cual le permite al usuario obtener modelos de predicción de producción y analizar el comportamiento de los sistemas de levantamiento artificial.

Esta primera versión esta creada para aplicar la metodología CRISP-DM a un conjunto de datos, la cual contempla: carga de datos, análisis inicial y comprensión de los datos, preparación y caracterización estadística de datos, modelado y por ultimo la evaluación del modelo.

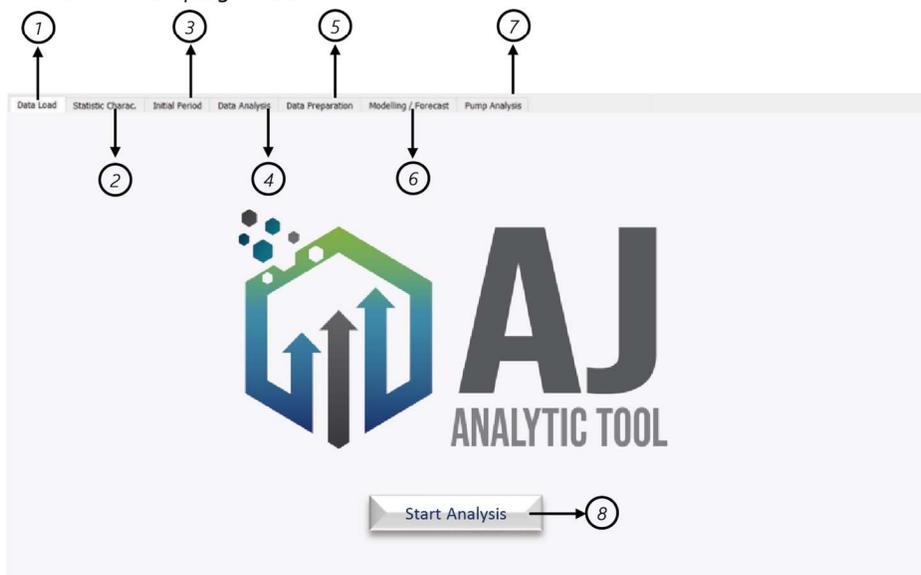
Para el modelado de predicción el software cuenta con 2 modelos de machine learning, (regresión lineal multivariable y regresión polinomial) y 1 modelo de deep learning (redes neuronales). Mientras, para el modelado de los sistemas de levantamiento el software usa el modelo de árbol de decisiones.

El siguiente manual muestra todos los procedimientos y algoritmos desarrollados y una breve explicación de las gráficas interactivas resultantes.

Elementos del programa

La interfaz cuenta con diferentes pestañas superiores que permiten al usuario desarrollar las diferentes tareas, cada una de las ventanas cuenta con botones programados específicos.

1. *Data load*: Pestaña que permite la carga de archivos a la herramienta
2. *Statistic Charac.*: Pestaña que muestra una descripción estadística de las bases de datos cargadas.
3. *Initial Period*: Pestaña que le permite al usuario analizar el periodo inicial antes de la producción.
4. *Data analysis*: Pestaña que permite al usuario analizar las bases de datos cargadas y seleccionar los rangos de cada parámetro.
5. *Data Preparation*: Pestaña que permite la preparación de los datos antes de la fase de modelado, incluyendo el procedimiento de caracterización estadística.
6. *Modelling/ Forecast*: Pestaña que permite al usuario realizar el proceso de modelado de los datos y obtener resultados de la predicción.
7. *Pump Analysis*: Pestaña que le permite al usuario realizar el análisis del funcionamiento de las bombas.
8. *Botones programados*.



Seleccionar archivos y carpetas

En esta etapa se seleccionan y cargan todos los archivos (.xlsx) y carpetas necesarias para realizar el análisis.

El programa requiere 4 archivos y 1 directorio para poder realizar el análisis de los datos del pozo.

Los mismos se detallan a continuación:

- **Data Master:** Archivo Excel que contiene la información de todos los pozos del campo en estudio. Esta información incluye campos como el nombre del pozo (WELLNAME), la macolla a la que pertenece (CLUSTER), el tipo de ambiente sedimentario (ASSET), la arena que produce (RESERVOIR), entre otros.
- **Production History:** Archivo Excel que contiene el historial de producción y operación de los pozos. Posee información como la fecha (DATE), el nombre del pozo (WELLNAME), las tasas de producción (EHO_RATE, WATER_RATE, GAS_RATE), velocidad de la bomba (RPM), entre otros.
- **Services:** Archivo Excel que contiene la información de los servicios y estimulaciones que se les han realizado a los pozos. Este documento debe contener: nombre del pozo (WELLNAME), tipo de servicio que se realizó (Event_Type), el modelo de la bomba si aplica (Pump_Model), entre otros.
- **Pumps Information:** Archivo Excel que contiene la información de catálogo de las bombas, como: modelo (MODEL), marca (BRAND) y desplazamiento nominal en bbl/rpm (DISPLACEMENT).
- **Plots Location:** Carpeta en donde se van a guardar los documentos y gráficos que se generan a lo largo del análisis.

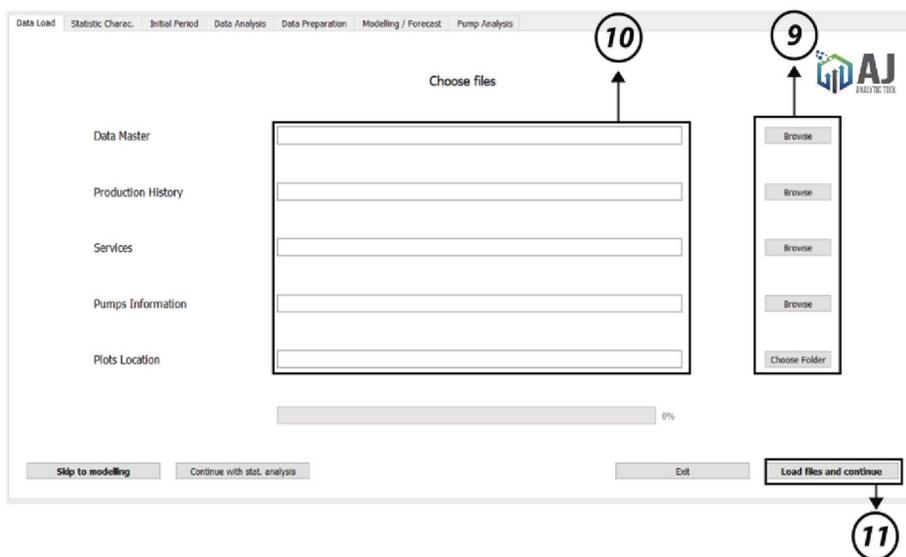
Seleccionar Archivos y carpetas

Para seleccionar los documentos y la carpeta descritos anteriormente, el usuario debe hacer clic en el botón **"Browse"** correspondiente al archivo que se desea cargar (9). Al hacer clic el programa mostrará una ventana de navegación, en la cual debe seleccionar el archivo deseado. Una vez que se seleccione el archivo, la ruta del mismo se mostrará en el campo de dirección correspondiente (10). También puede escribir la ruta del archivo en el campo de dirección directamente.

Nota: A lo largo del proceso de análisis, el programa generará hojas de Excel (.xlsx) y gráficos en formato HTML (.html) que se guardarán en la carpeta seleccionada.

Para avanzar a la siguiente etapa debe cargar todos los archivos y seleccionar la ruta directorio, posteriormente el usuario debe hacer clic en el botón **"Load files and continue"** (11).

Todos los botones se desactivarán y aparecerá una barra de estado que refleja el progreso de la carga de los datos. Al finalizar el proceso, el sistema emite un sonido breve que indica que los documentos han sido cargados exitosamente.



Selección y análisis de los pozos cargados

En esta etapa, se selecciona el pozo a analizar y se muestran algunas características de interés de los datos cargados.

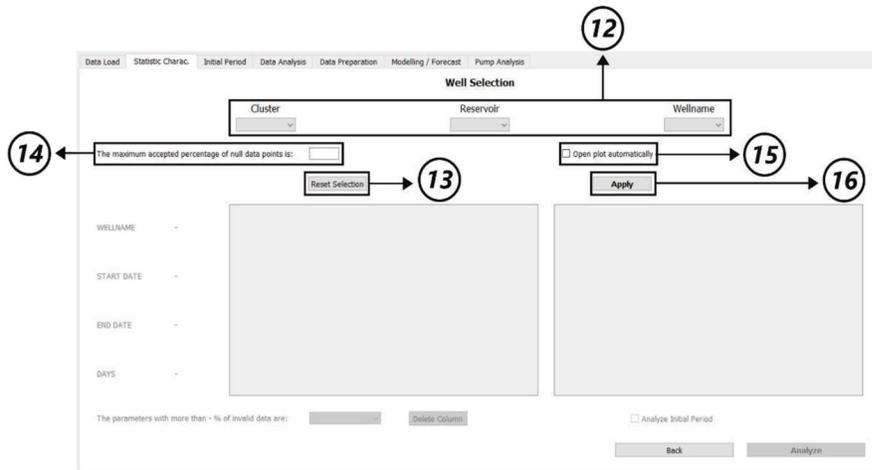
Para seleccionar un pozo, el usuario debe seleccionar en las listas desplegables (12) la macolla (Cluster), la arena (Reservoir) y el nombre del pozo (Wellname).

Nota: Al seleccionar alguno de los parámetros, los otros dos se filtran automáticamente, mostrando solo aquellos ítems que pertenecen a dicha selección. Si ya se ha seleccionado el pozo y se desea cambiar la selección, hacer clic en el botón **"Reset Selection"** (13) para reiniciar las listas desplegables.

Una vez que haya seleccionado el pozo, el usuario debe colocar el porcentaje máximo inicial (sólo valores enteros) de datos nulos/vacíos que se aceptará en cada uno de los parámetros en estudio. Este valor debe ingresarlo en la casilla de entrada del porcentaje (14) de los parámetros en estudio. **Nota:** El valor preestablecido es 10%.

La casilla de verificación **"Open plot automatically"**, abre un gráfico inicial de todos los parámetros cargados (15). Si la casilla del gráfico está seleccionada, al momento de hacer clic en el botón **"Apply"**, el navegador abrirá automáticamente el gráfico interactivo (las funcionalidades del gráfico se explicarán más adelante en este manual).

Al completar los pasos anteriores, hacer clic en el botón **"Apply"** (16). Este botón desactivará los elementos anteriores, y aparecerá una ventana emergente con un mensaje. Hacer clic en OK, y esperar hasta que salga la información en pantalla.



Análisis de los datos por pozo

Una vez que los datos son procesados, en la parte izquierda de la ventana (17), se visualizará la información sobre el pozo seleccionado: nombre (WELLNAME), fecha de inicio de operación (START DATE), fecha final de operación (END DATE) y los días totales transcurridos entre el inicio y el fin de las operaciones (DAYS).

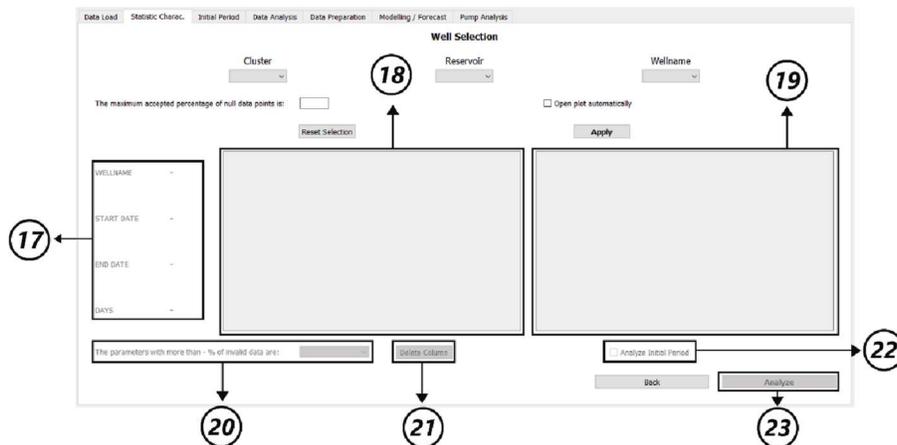
En la parte central de la ventana (18) se mostrarán los parámetros que se encuentran en los datos cargados, la cantidad de datos nulos/vacíos que hay en cada parámetro y el porcentaje que éstos representan respecto al total.

En la parte derecha de la ventana (19) se muestra un resumen estadístico de cada uno de los parámetros que se cargaron durante la primera etapa. La información que se muestra es la siguiente:

- **Count:** cantidad de datos totales que hay en cada uno de los parámetros.
- **Mean:** promedio de todos los datos numéricos que hay en determinado parámetro.
- **Std:** desviación estándar del conjunto de valores.
- **Min:** valor mínimo que existe en todo el rango.
- **25%:** valor del percentil 25 de los datos.
- **50%:** valor del percentil 50 de todos los datos (mediana).
- **70%:** valor del percentil 75 de todos los datos.
- **Max:** valor máximo que existe en todo el rango.

En caso de que en alguno de los parámetros exista un porcentaje de datos nulos mayor al establecido en el paso anterior, dicho parámetro aparecerá en la lista desplegable de la parte inferior (20).

Si se desea eliminar ese parámetro completamente, debe seleccionarlo de la lista y luego hacer clic en el botón **"Delete Column"** (21).



Análisis del periodo inicial en el pozo en estudio

Nota: Si se desea analizar los valores del periodo inicial de la vida del pozo (desde el primer día de la toma de datos hasta el primer día de funcionamiento de la bomba), activar la casilla *"Analyze initial Period"* (22).

Una vez que se hayan eliminado las columnas deseadas, hacer clic en el botón *"Analyze"* (23).

Si la casilla *"Analyze initial Period"* está seleccionada, se activarán dos pestañas: *"Initial Period"* y *"Data analysis"*.

- En la pestaña *"Initial period"* aparece una ventana única (24), con datos estadísticos importantes como el promedio, el valor mínimo, el valor máximo, entre otros.

24

Attribute	count	mean	std	min	25%	50%	75%	max
DURATION	104.00	1.33	4.25	0.00	0.00	0.00	0.00	20.95
WHP	103.00	28.04	21.84	14.70	14.70	14.70	43.69	126.36
WHT	103.00	88.17	3.02	79.44	86.44	87.89	89.42	96.84
RPM	104.00	5.21	15.24	0.00	0.00	0.00	0.00	50.84
TORQUE	103.00	73.58	170.87	0.00	0.00	0.00	0.00	650.33
INTAKE_PRESS	104.00	601.40	112.41	0.00	619.99	630.20	631.99	632.96
DISCHARGE_PR...	103.00	-2198.04	20194.53	-143209.00	626.39	629.71	630.63	760.37
DISCHARGE_TE...	103.00	113.83	14.33	82.00	116.16	116.96	117.34	120.96
DILUENT_VOL	104.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GAS_RATE	104.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
WATER_RATE	104.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PI	104.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PUMP_EFFICIENC	104.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SHO_RATE	104.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Category	104.00	0.06	0.58	0.00	0.00	0.00	0.00	6.00
Pump_Displace...	104.00	5.16	0.00	5.16	5.16	5.16	5.16	5.16
LIQUID_RATE	104.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
t	104.00	51.50	30.17	0.00	25.75	51.50	77.25	103.00

Análisis de los datos

Esta ventana muestra los rangos mínimos y máximos de cada uno de los parámetros (25) así como la cantidad de datos inválidos/nulos/vacíos en cada uno (26).

En esta etapa, se procede a modificar los valores mínimos y máximos de los rangos considerados como aceptables a juicio del usuario.

A medida que se van modificando los valores límites, la cantidad de datos inválidos/fuera del nuevo rango se actualizará automáticamente.

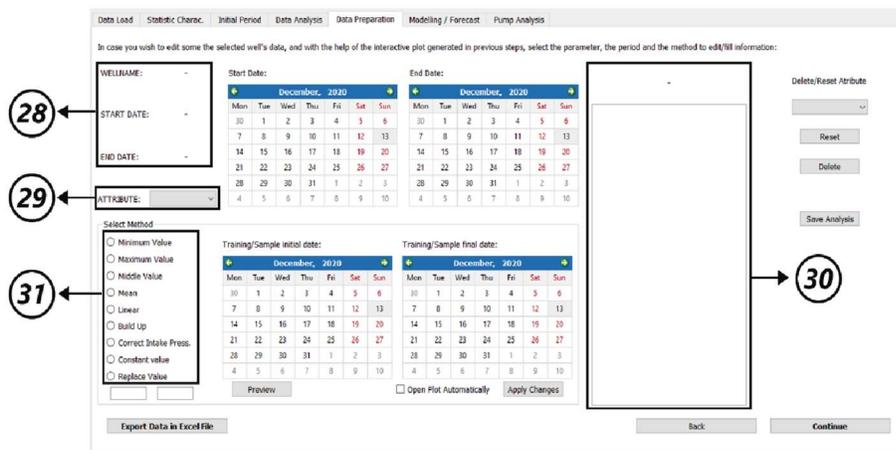
Al terminar de corregir los rangos, el usuario debe hacer clic en el botón “Continue” (27). Este botón abrirá automáticamente un gráfico interactivo y activará la pestaña “Data Preparation”.

Preparación de los datos | Caracterización estadística

En esta etapa se realizan todos los cambios necesarios a los datos, con el fin de corregir errores y prepararlos para el entrenamiento de los modelos de machine learning y deep learning.

En el lado superior izquierdo (28), se muestra el nombre del pozo, la fecha de inicio y la fecha final de la vida productiva del pozo (de acuerdo a los datos cargados).

Para comenzar con la modificación de los datos, primero, el usuario debe hacer clic en la lista desplegable (29) para seleccionar el parámetro (ATTRIBUTE) que se va a editar, una vez que se selecciona el parámetro, en la sección (30) verá el nombre del parámetro que se está editando, y todos los datos que se consideran erróneos, con sus respectivas fechas.



Una vez que se hayan escogido las fechas y valores que se van a corregir, lo primero que hay que hacer es seleccionar el método de corrección deseado entre las opciones presentadas (31). Entre las métodos disponibles están: minimum value, maximum value, middle value, mean, linear, build up, correct intake press., constant value y replace value.

Preparación de los datos | Caracterización estadística

Dependiendo de la opción escogida, se mostrarán u ocultarán algunos campos necesarios para poder ingresar la información y realizar la corrección de los datos. Los campos para ingresar información son los siguientes:

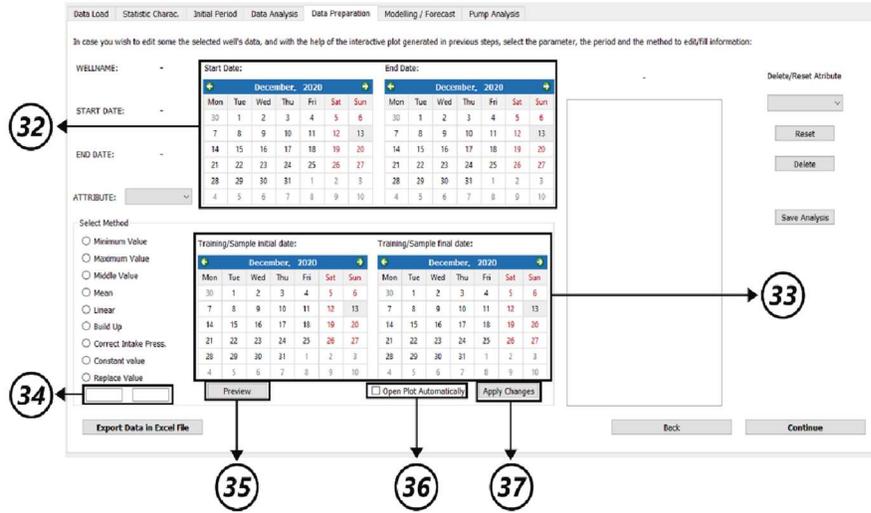
- **Calendarios Start Date y End Date (32):** Son dos calendarios que se muestran en la parte central superior de la pantalla, dichos campos sirven para poder ingresar el periodo de tiempo que el usuario desea corregir. En Start Date se selecciona el primer día del periodo que se desea corregir, mientras que en End Date se selecciona el último día. El calendario Start Date aparece para todas las opciones, mientras que el calendario End Date aparece para todas las opciones a excepción de la opción Midde Value.
- **Calendarios Training/Sample initial date y Training/Sample final date (33):** Estos calendarios sirven para seleccionar un periodo de tiempo en el que el parámetro seleccionado tenga un comportamiento deseado, y que se desee replicar en el periodo que se va a corregir (métodos de Build Up y Correct Intake Press). También se usa para seleccionar un periodo en el que se desee obtener el valor mínimo, el valor máximo o el valor promedio para luego aplicarlo al periodo que se va a corregir (método Minimum Value, Maximum Value y Mean).

Nota: En los calendarios, la fecha mínima que se puede seleccionar es el primer día de producción del pozo, y la última fecha es el último día del historial de producción de dicho pozo. Cada vez que se seleccione una fecha en Start Date o en Training/Sample initial Date, automáticamente se seleccionará la misma fecha en End Date o Training/Sample final date. Adicionalmente, en el campo End Date no se pueden seleccionar fechas anteriores a la fecha seleccionada en Start Date.

- **Cuadro(s) de texto (34):** se usan para ingresar uno o dos valores numéricos, dependiendo de la opción escogida. Si se seleccionó la opción Constant Value, sólo aparece un cuadro, y si se seleccionó Replace Value se muestran los dos. En el cuadro donde aparece el texto Old Val se coloca el valor numérico que se desea reemplazar, y en el cuadro donde aparece New Val, se coloca el valor por el cual se desea reemplazar.
- **Botón Preview (35):** Este botón se usa para poder visualizar en un gráfico como se verían los cambios que se desean aplicar, antes de aplicarlos sobre los datos, de esta manera se puede observar si dichas correcciones tienen el efecto deseado. Este botón se muestra para todas las opciones.
- **Casilla Open Plot Automatically (36):** es una casilla de verificación que, si se encuentra activa al momento de hacer clic en el botón "*Apply changes*", abre en el navegador un gráfico en donde se muestran todos los parámetros en estudio incluyendo las correcciones realizadas. Esta casilla se muestra para todas las opciones.

Preparación de los datos | Caracterización estadística

Botón *Apply changes* (37): Este botón se usa para aplicar las correcciones seleccionadas. Aparece para todas las opciones.



Si en algún momento, se desea eliminar un atributo completamente o se desea restaurar los valores a los iniciales (incluidos todos sus valores erróneos), primero hacer clic en la lista desplegable que se encuentra en la parte derecha (38), y seleccionar el atributo deseado, una vez que se haya seleccionado, se activaran dos botones por debajo de la lista: **"Reset"** y **"Delete"** (39). El botón **"Reset"** sustituye todos los valores de ese parámetro por los valores iniciales, por lo que todos los cambios hechos hasta el momento quedan sin efecto, mientras que el botón **"Delete"** elimina el atributo seleccionado de los datos.

Preparación de los datos | Caracterización estadística

Si se desea guardar los datos con las correcciones realizadas hasta dicho momento, hacer clic en el botón **"Save Analysis"** (40). Esto creará tres archivos .csv con el estado de los datos hasta ese momento, e información adicional necesaria para retomar el análisis más adelante.

Una vez que se hayan realizado todas las correcciones necesarias, hacer clic en el botón **"Export Data in Excel File"** (41), para crear un archivo Excel (.xlsx) con toda la información ya procesada.

Cuando se haya finalizado el pre-procesamiento de los datos, hacer clic en el botón **"Continue"** (42), para proceder a la etapa de modelado y predicción (Modelling / Forecast).

In case you wish to edit some the selected well's data, and with the help of the interactive plot generated in previous steps, select the parameter, the period and the method to edit information:

WELLNAME: -

START DATE: -

END DATE: -

ATTRIBUTE: -

-Select Method

- Minimum Value
- Maximum Value
- Middle Value
- Mean
- Linear
- Build Up
- Correct Intake Press.
- Constant value
- Replace Value

Buttons: Export Data in Excel File (41), Save Analysis (40), Continue (42), Back, Reset (39), Delete (39), Back (38).

Modelado de los datos

En esta ventana, en la parte superior derecha, hay un mensaje y dos botones (43). Si se desea entrenar un modelo de Machine Learning para posteriormente realizar una estimación de la producción basado en dicho modelo, hacer clic en el botón *"Model training"*. Si por el contrario, ya se tiene un modelo previamente definido y se desea hacer la estimación directamente, hacer clic en el botón *"Forecast"*.

Model Training / Entrenamiento del modelo:

Al seleccionar la opción *"Model training"*, se va a activar la sección superior izquierda *"Select training Method"*, en donde se tiene que escoger una de las cuatro opciones disponibles (44):

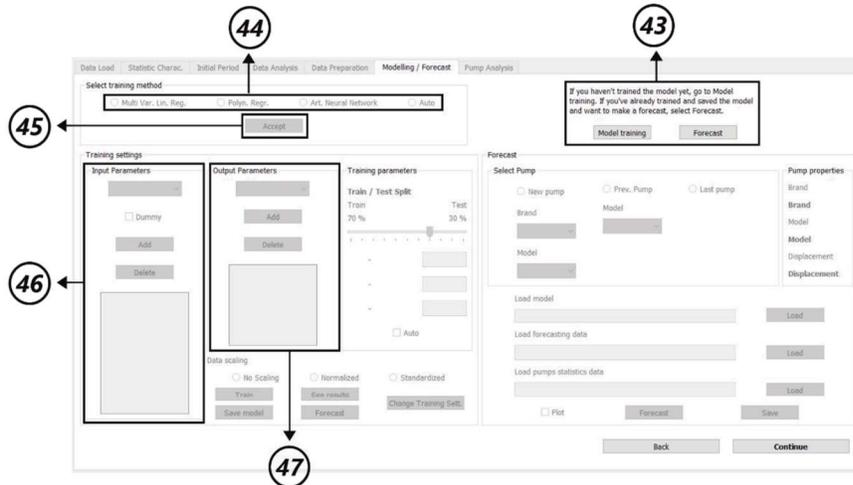
- Regresión Lineal Multivariable (Multivariable Linear Regression)
- Regresión Polinomial (Polynomial Regression)
- Redes neuronales artificiales (Artificial Neural Network)
- Automático (Auto): aplica las tres metodologías mencionadas anteriormente y selecciona la que obtenga mejores resultados, es decir, un menor error relativo entre las estimaciones y los valores reales de los datos de prueba.

Una vez que se haya seleccionado el modelo que se va a entrenar, presionar el botón *"Accept"* (45). Dependiendo de la opción seleccionada, se van a activar/desactivar algunas opciones para ajustar los parámetros del modelo. Entre las opciones se encuentran:

Input Parameters (Parámetros de entrada) (46): Son aquellos datos que van a ser las variables de entrada (independientes) del modelo (tomar en cuenta que, para hacer las predicciones, es necesario ingresar los valores estimados de estas variables). Dependiendo del modelo escogido, van a aparecer pre-seleccionados algunos campos que son necesarios para el código interno del modelo escogido. Para seleccionar campos adicionales, hacer clic en la lista desplegable que se encuentra en la parte superior, seleccionar el parámetro deseado y hacer clic en el botón *"Add"*. Si el parámetro deseado es una variable no numérica, es necesario transformarla a variable tipo numérica. Para ello, al seleccionar dicha variable (por ejemplo, Pump Model), marcar la casilla *"Dummy"*, y luego hacer clic en *"Add"*. Si se desea eliminar alguna de las variables ya escogidas, seleccionarla de la lista inferior, y hacer clic en el botón *"Delete"*. Esto hará que dicha variable aparezca nuevamente en la lista desplegable.

Modelado de los datos

Output Parameters (Parámetros de salida) (47): Es aquel parámetro que se desea predecir/estimar (variable dependiente), solo se permite uno. Cuando se selecciona la opción auto, la variable predeterminada es EHO_RATE (tasa de petróleo). Se aplican los mismos controles explicados en Input Parameters.



Train/test Split (división Entrenamiento/Prueba) (48): Corresponde a la división de los datos en dos partes, una de entrenamiento y otra de prueba. Para ajustar estos porcentajes, deslizar el control del deslizador, y al hacerlo, se indicarán los valores, tanto para el entrenamiento como para la prueba. Los valores predeterminados (y los más usados para Machine Learning) son 70% de entrenamiento y 30% de prueba.

Degree (Grado) (49): Es el grado del polinomio resultante de la regresión polinomial. Este campo se activa solo cuando el método seleccionado es la Regresión Polinomial. Para ingresar el grado deseado, escribirlo en la casilla. Tomar en cuenta que, a mayor grado, se requiere un mayor tiempo y mayor poder computacional para realizar el modelo.

Neurons (Neuronas) (49): Este campo es el mismo campo de "Degree". Se activa solamente cuando el método seleccionado es "Artificial Neural Network" (Redes Neuronales Artificiales). Representa la cantidad de neuronas en cada una de las capas intermedias de la red neuronal.

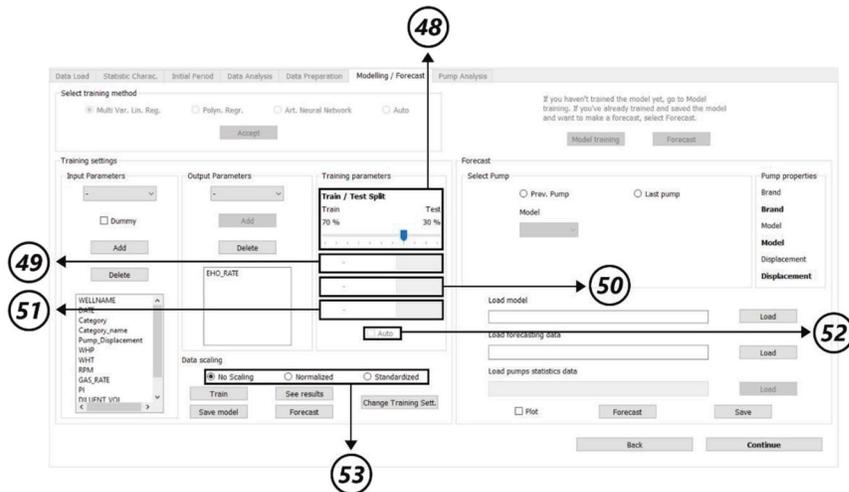
Modelado de los datos

Layers (Capas) (50): Este campo se activa solamente cuando el método seleccionado es *"Artificial Neural Network"* (Redes Neuronales Artificiales). Representa la cantidad de capas intermedias (también llamadas Ocultas) de la red Neuronal.

Epochs (Iteraciones) (51): Este campo se activa solamente cuando el método seleccionado es *"Artificial Neural Network"* (Redes Neuronales Artificiales). Representa la cantidad de iteraciones o veces que la red neuronal calcula la propagación de los errores para minimizarlos.

Auto (52): Esta casilla se activa solamente cuando el método seleccionado es *Artificial Neural Network* (Redes Neuronales Artificiales). Al marcar esta casilla, se desactivan las tres casillas anteriores, ya que internamente, el programa va a entrenar varias redes neuronales con configuraciones predeterminadas, y va a seleccionar el que presente un menor error.

Data Scaling (Escala de los datos) (53): En esta sección, se selecciona una de las opciones para realizar una escala de los datos antes del entrenamiento, las opciones disponibles son *"No Scaling"* (sin escala), *"Normalized"* (Normalizados, es decir, los valores varían entre 0 y 1) y *"Standardize"* (Estandarizados).



Modelado de los datos / Predicción de producción

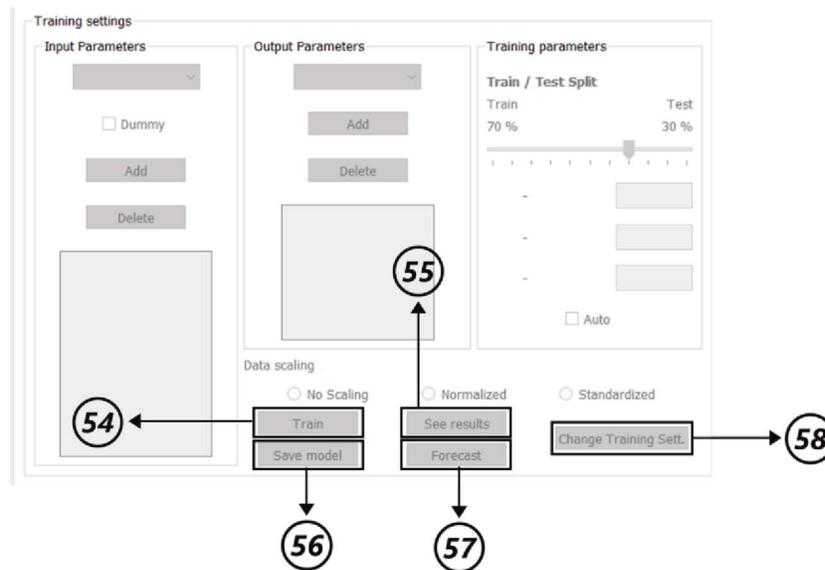
Una vez que se hayan ajustado los parámetros de entrenamiento, hacer clic en el botón **"Train"** (54), y esperar que se termine el proceso. Una vez que el entrenamiento haya finalizado, se activaran los otros botones de ésta sección.

See Results (55): Para abrir un documento de Excel mostrando los errores y resultados. Adicionalmente, se abrirá un gráfico en formato HTML, donde se muestra los valores reales de la variable dependiente, y los valores estimados de acuerdo al modelo.

Save Model (56): Guarda un archivo con extensión .pickle o .h5 (dependiendo del modelo seleccionado), con los parámetros y valores del modelo, en caso de que en otra sesión se desee estimar la producción usando dicho modelo.

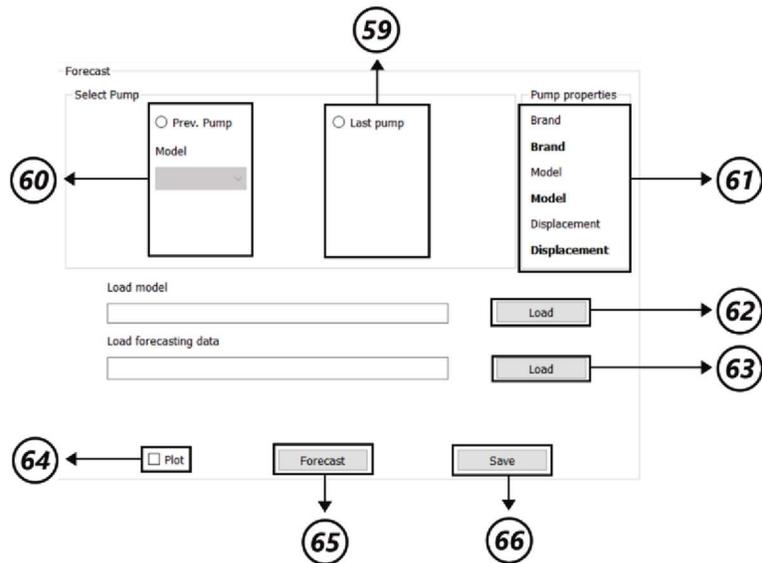
Forecast (57): Esta opción permite activar la sección de **"Forecast"**, en donde se configuran las opciones para realizar una estimación de la producción.

Change Training Settings (58): Esta opción permite cambiar las configuraciones y volver a entrenar un nuevo modelo, usando los datos ya cargados.



Predicción de producción

Para hacer una predicción / estimación usando algún modelo generado previamente, primero hay que seleccionar la bomba que estaría trabajando en el periodo que se desea estimar. Para su selección hay que elegir una de las dos opciones presentadas: la última bomba usada (Last Pump) (59) o alguna de las bombas usadas con anterioridad (Prev. Pump) (60), estos modelos de bombas dependen de los datos usados para el entrenamiento del modelo. Una vez que se haya seleccionado la bomba, sus datos (Marca, modelo y desplazamiento nominal de fluido) van a aparecer en la parte derecha (61). Luego, hay que seleccionar 2 archivos necesarios para realizar la estimación: el archivo que contiene el modelo previamente entrenado (extensión .pickle o .h5) (62), y el documento Excel que contiene los valores estimados de los parámetros independientes (dicho documento fue generado al momento de entrenar y guardar el modelo) (63). Si se desea visualizar un gráfico donde se aprecie la estimación de tasa en función del tiempo, marcar la casilla "Plot" (64). Una vez que se hayan seleccionado los parámetros necesarios, hacer clic en el botón "Forecast" (65) para estimar la tasa y mostrar el gráfico, si es que se seleccionó la opción correspondiente. Para guardar un documento Excel con los valores, tanto de los parámetros independientes como las tasas estimadas, hacer clic en el botón "Save" (66).



Análisis de bombas

Nota: Para realizar un análisis del desempeño de las bombas en un determinado pozo durante el historial de producción, se necesita haber realizado previamente, el modelado del pozo por el método de regresión lineal, ya que, al realizar este paso, se generará un documento Excel necesario para este análisis.

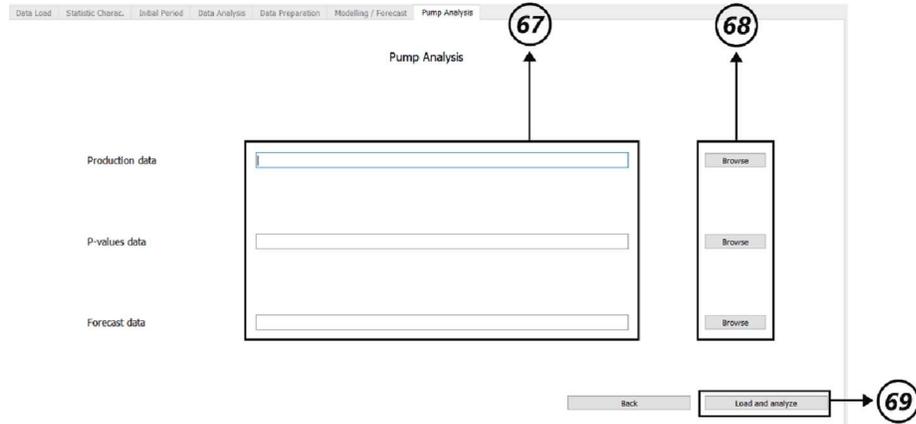
En la pantalla para el análisis de las bombas se requieren 3 documentos:

Production data: este archivo es el documento Excel que contiene el historial de producción del pozo en estudio.

P-Values Data: este archivo es el documento Excel que fue generado por el programa cuando se entrenó el modelo de regresión lineal Multivariable, y que contiene los P-values de cada una de las variables independientes usados para generar el modelo. Generalmente tienen en el nombre la frase "Result of linear regression"

Forecast data: es el documento Excel generado por el programa como resultado de entrenar uno de los modelos con los datos del pozo. Si el modelo seleccionado es el de regresión lineal Multivariable, este archivo es el mismo que el archivo anterior. Para seleccionar los documentos, ingresar la ruta de documento requerido en el campo correspondiente (67), o haciendo clic en el botón correspondiente (68) y seleccionarlo con la ayuda del cuadro de dialogo.

Una vez que se hayan seleccionado los 3 documentos, hacer clic en el botón "Load and Analyze" (69) y esperar que la información se haya procesado.



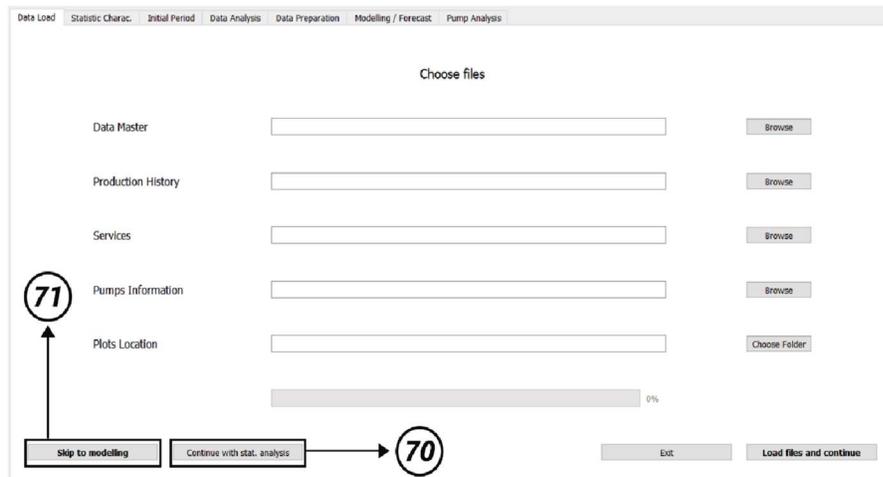
Retomar el trabajo

Opción 1: Continuar con la caracterización estadística de los datos

Si por alguna razón no se puede terminar el análisis estadístico de los datos (sección Data Preparation), en la pantalla inicial (Data Load), seleccionar la opción **"Continue with stat. Analysis"** (70). Al hacer esto, se abrirán 3 ventanas emergentes consecutivas, en las cuales se seleccionarán 3 documentos de tipo .csv: los datos parcialmente caracterizados (**incomplete stat. Analysis**), los rangos aceptables de cada uno de los parámetros (**Ranges**) y los datos originales (**Original values**). Una vez que se hayan cargado dichos documentos, se activará la pestaña **"Data Preparation"**, y así poder continuar con el análisis de los datos.

Opción 2: Proceder al entrenamiento de los modelos del pozo

Si ya se tienen todos los datos ya caracterizados y corregidos, y se desea saltar directamente al entrenamiento del modelo usando alguna metodología de Machine Learning, hacer clic en el botón **"Skip to modelling"** (71). Al hacer esto, se abrirán dos ventanas emergentes consecutivas, en las cuales se seleccionarán dos archivos: los datos ya caracterizados y corregidos del pozo que se va a analizar (un documento con extensión .xlsx) y una carpeta que es donde se van a guardar todos los archivos que se generarán a lo largo del proceso. Una vez que se hayan seleccionado dichos archivos, se activará la pestaña **"Modelling/Forecast"**, y así poder dar inicio al entrenamiento del modelo.



Funciones de los gráficos interactivos

A lo largo del procedimiento de análisis usando el programa, en algunos pasos se van a generar gráficos interactivos en formato HTML. Algunas veces el programa va a abrir dichos gráficos de manera automáticamente, lo cual se va a hacer usando el navegador predeterminado (Mozilla Firefox, Google Chrome, etc.) y otras veces simplemente se va a guardar dicho archivo en la carpeta seleccionada sin abrir ningún navegador.

Cualquiera sea el caso, los gráficos generados van a tener ciertas funciones para poder visualizar mejor los trazos y que el usuario pueda interactuar con él. A continuación, se muestran las funcionalidades:

Cuando el mouse está sobre el gráfico, en la parte superior derecha se muestra una barra de herramientas (A) con 10 funcionalidades que son

Download plot as png: Descargar grafico como archivo png

Zoom: Seleccionar un área del gráfico y realizar un acercamiento a dicha área.

Pan: Moverse a través del grafico al arrastrar el mouse sobre el grafico.

Zoom in: Acercar

Zoom out: Alejar

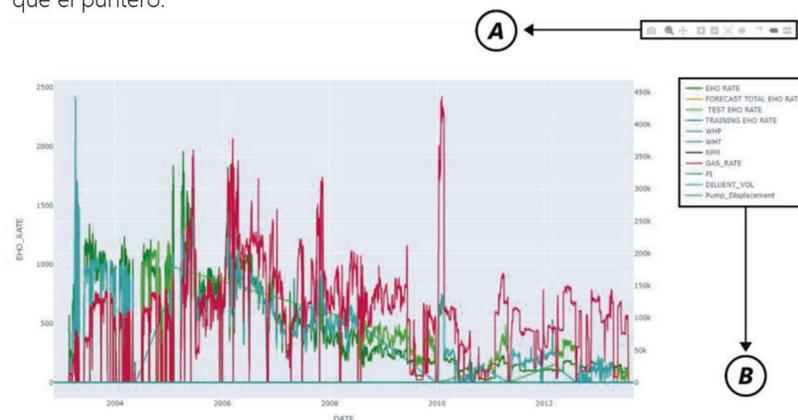
Autoscale: Autoajustar la escala del gráfico.

Reset axes: Reiniciar los valores de los ejes

Toggle spike lines: Mostrar líneas paralelas a los ejes que muestren las coordenadas de un punto del grafico

Show closest data on hover: al acercar el puntero del mouse a la gráfica, se muestra una etiqueta con los datos del punto del grafico más cercano al puntero.

Compare data on hover: al mover el puntero del mouse sobre el grafico, se muestran los valores de todos los parámetros que están en la misma coordenada X que el puntero.

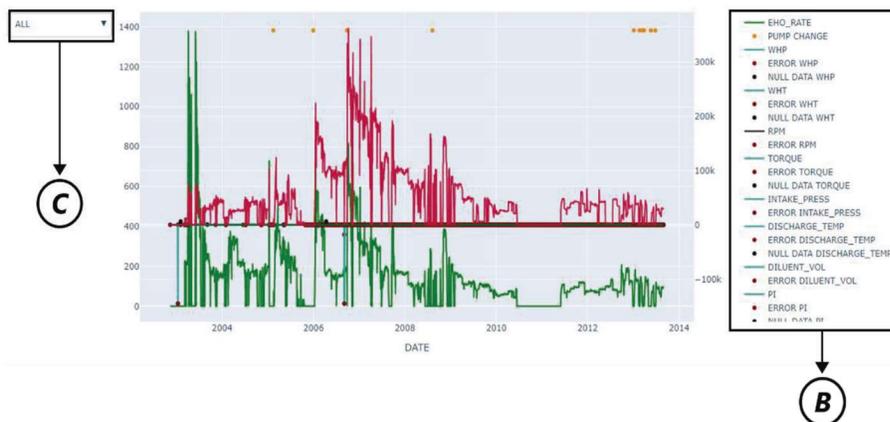


Funciones de los gráficos interactivos

Adicionalmente a los botones de la mencionada barra de herramientas, la leyenda del gráfico (B) también es interactiva, ya que, al hacer clic sobre alguno de los valores, el trazo correspondiente se activa/desactiva en el gráfico. Si se hace doble clic sobre algún valor de la leyenda, todos los demás trazos desaparecen del gráfico, a excepción del valor seleccionado.

Algunos de los gráficos interactivos tienen una lista desplegable en el lado izquierdo con todos los parámetros que se deseen estudiar (C). Al seleccionar un parámetro, tanto en el gráfico como en la leyenda se muestran la tasa de petróleo (EHO_RATE), los cambios de bomba (PUMP CHANGE), y los trazos y puntos relacionados a ese parámetro tales como datos nulos/vacíos (NULL DATA) y datos fuera del rango establecido previamente (ERROR).

Todas las gráficas interactivas que se generan a lo largo del análisis, tienen las mismas funcionalidades básicas que se describen aquí.



<i>Elementos del programa</i>	3
<i>Data load</i>		
<i>Seleccionar archivos y carpetas</i>	4
<i>Statistic. Charac.</i>		
<i>Selección y análisis de los pozos cargados</i>	6
<i>Análisis de los datos por pozo</i>	7
<i>Initial Period</i>		
<i>Análisis del periodo inicial en el pozo en estudio</i>	8
<i>Data Analysis</i>		
<i>Análisis de los datos</i>	9
<i>Data Preparation</i>		
<i>Preparación de los datos / Caracterización estadística</i>	10
<i>Retomar el trabajo</i>	20
<i>Modelling / Forecast</i>		
<i>Modelado de los datos</i>	14
<i>Predicción de producción</i>	17
<i>Análisis de bombas</i>	19
<i>Gráficos interactivos</i>	21

BIBLIOGRAFÍA

1. *Propiedades de los fluidos*. Padrón, Alida. Caracas : s.n., 2017. pág. 29.
2. Alida, Padrón. *Propiedades de los Yacimientos*. [Presentación] Caracas : s.n., 2014.
3. Artículos técnicos. Crudos pesados y extrapesados. *La comunidad petrolera*. [En línea] 24 de Octubre de 2008. [Citado el: 5 de marzo de 2020.] <https://www.lacomunidadpetrolera.com/2008/10/crudos-pesados-y-extrapesados.html>.
4. *Importancia del proceso de producción*. Miranda, Lisbeth. Caracas : s.n., 2018. pág. 10.
5. Compresibilidad de la roca. *PerfoBlogger*. [En línea] 14 de Febrero de 2016. <http://perfob.blogspot.com/2016/02/compresibilidad-de-la-roca.html>.
6. Schlumberger. Oilfield Glossary. [En línea] 2019. [Citado el: 17 de 12 de 2019.] <https://www.glossary.oilfield.slb.com/es/Terms/c/consolidated.aspx>.
7. Carita, Gabriela. *¿Cuál es la diferencia entre los sedimentos consolidados y no consolidados?* [En línea] 16 de Febrero de 2016. [Citado el: 17 de 12 de 2019.] <http://gidahatari.com/ih-es/cual-es-la-diferencia-entre-los-sedimentos-consolidados-y-no-consolidados>.
8. Horne., Roland N. *Modern Well test analysis – A computer aided approach (1ra edición)*. 1990.
9. Petrocedeño. *Principios Básicos de Producción*. [Presentación Power Point] Caracas : s.n., 2010.
10. Statoil. *Procesos básicos de producción*. [Presentación PowerPoint] Caracas : s.n., 2010.
11. Díaz, Pedro. *Guía de Pozos I*. [Guía impresa No publicada] Caracas : s.n., 2018.
12. Schlumberger Glossary. Presión estática. *Glossary Oilfield Schlumberger*. [En línea] 2019. [Citado el: 19 de 12 de 2019.] https://www.glossary.oilfield.slb.com/es/Terms/s/static_pressure.aspx.
13. Chacin, Nelvy. *Bombeo de Cavidad Progresiva - ESP Oil International Training Group*. [Documento PDF] 2003.

14. Alvarado, Eliana. *Clases Yacimientos III*. [Presentación PowerPoint] Caracas : s.n., 2018.
15. Hernandez, Edgar. *Clases Análisis de Pruebas de Presión*. [Presentación PowerPoint] Caracas : s.n., 2018.
16. Petrowiki. *Well test*. [En línea] [Citado el: 19 de Febrero de 2020.] https://petrowiki.org/Well_test.
17. Productivity index. *Petrowiki*. [En línea] [https://petrowiki.org/Productivity_index_\(J\)](https://petrowiki.org/Productivity_index_(J)).
18. Equipetrol. Equipetrol. *Valvulas Seleccion Multipuerto*. [En línea] [Citado el: 2020 de 2 de 15.] <https://equipetrol.us/es/valvulas-selectoa-multipuerto/>.
19. Oilfield Glossary Schlumberger. Oilfield Glossary Medidor multifase. *Schlumberger*. [En línea] [Citado el: 19 de Febrero de 2020.] https://www.glossary.oilfield.slb.com/es/Terms/m/multiphase_meter.aspx.
20. Diccionario petrolero. La comunidad petrolera. *Bombeo multifásico*. [En línea] 24 de Junio de 2009. [Citado el: 15 de Febrero de 2020.] <https://www.lacomunidadpetrolera.com/2009/06/bombeo-multifasico.html>.
21. Essenfeld, Martin. *Yacimientos de Hidrocarburos*. Caracas : FONCIED, 2001.
22. Jaramillo, Edgar Ramirez. *Estudio del problema de arenamiento en pozos productores de gas, propuesta de modelo*. Ciudad de México : s.n., 2016.
23. Estimulación. *Schlumberger Oilfield Glossary*. [En línea] <https://www.glossary.oilfield.slb.com/es/Terms/s/stimulation.aspx>.
24. Hirschfeldt, Marcelo. *Manual de bombeo de cavidades progresivas* . 2008.
25. Aranda, Felipe de Jesus Lucero. *Manual del participante - Especialidad en ingeniería petrolera*. . Mexico : s.n., 2008.
26. Protex Complete Integrated Solutions. *Progressive Cavity pump manual (version 07/03)*. 2007.
27. BCPVEN. *Bombas de cavidad progresiva BCP*. [En línea] [Citado el: 16 de Enero de 2020.] <http://www.bcpven.com/bomba-de-cavidad-progresiva-bcp/>.

28. Petrowiki. *Alternate PCP system configurations*. [En línea] [Citado el: 29 de Enero de 2020.] https://petrowiki.org/Alternate_PCP_system_configurations#Charge_pump_systems.
29. Mono, Nov. *Progressing Cavity Pump Systems*. 2013.
30. Netzsch. *Manual de sistemas PCP*.
31. 2. Oferta y demanda de energía: tendencias y perspectivas. [En línea] [Citado el: 17 de marzo de 2020.] <http://www.fao.org/tempref/docrep/fao/010/i0139s/i0139s03.pdf>.
32. BP Statistical Review of World Energy 2019. *BP*. [En línea] 2019. [Citado el: 17 de marzo de 2020.] <https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2019-full-report.pdf>.
33. Alboudwarej, Hussein, Joao, Felix y Shawn, Taylor. La importancia del petróleo. *Schlumberger*. [En línea] 2006. [Citado el: 17 de marzo de 2020.] <https://www.slb.com/-/media/files/oilfield-review/heavy-oil-3-spanish>.
34. Press, Gil. Forbes. *A Very Short History Of Data Science*. [En línea] 28 de Mayo de 2013. <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#7efd3a8e55cf>.
35. Ortiz, Angel. Pc Web Info. *Historia del data science, la ciencia de datos*. [En línea] 11 de Enero de 2019. <https://pcweb.info/historia-del-data-science-la-ciencia-de-datos/>.
36. Manfre, Diego. *Análisis de datos para ingeniería de petróleo*. Caracas, Venezuela : s.n., 24 de Enero de 2020.
37. *Devon Enegy Rises to the top as a Data-Driven Producer*. Jacobs, Trent. 2016, Journal of Petroleum Technology, págs. 28-29.
38. *Functional approach to Data Mining, forecasting, and uncertainly quantification*. Ognen Grujic, Carla Da Silva, Jef Caers. 2016, Journal of Petroleum Technology, págs. 67-68.
39. Ognjen Grujic, Carla Da Silva, Jef Caers. *Data Driven production forecasting using machine learning*. s.l. : SPE 180984-MS, 2016.

40. *Big Data Analytics for predictive maintenance modeling: challenges and opportunities*. Santos, et al. 2016, Journal of Petroleum Technology, págs. 71-72.
41. Economipedia. *Estadística*. [En línea] [Citado el: 17 de Julio de 2020.] <https://economipedia.com/definiciones/estadistica.html>.
42. Neoland. Neoland. *¿Qué es Data Science?* [En línea] 12 de Marzo de 2019. [Citado el: 13 de Febrero de 2020.] <https://www.neoland.es/blog/que-es-data-science>.
43. Obiols, Albert. In Lab Talent Tech. *¿Qué es un Data Scientist?* [En línea] 20 de Mayo de 2015. [Citado el: 12 de Febrero de 2020.]
44. Ana Azevedo, Manuel Filipe Santos. *KDD, SEMMA and CRISP-DM: a parallel overview*. 2008.
45. Julio Villena Román. Singular. *CRISP-DM: La metodología para poner orden en los proyectos*. [En línea] 2 de Agosto de 2016. [Citado el: 5 de Enero de 2020.] <https://www.singular.com/es/data-science-crisp-dm-metodologia/>.
46. Wikipedia. *Metodología Semma*. [En línea] [Citado el: 10 de Enero de 2020.] <https://en.wikipedia.org/wiki/SEMMA>.
47. Smart Vision. *what is the CRISP-DM methodology?* [En línea] [Citado el: 10 de Enero de 2020.] <https://www.sv-europe.com/crisp-dm-methodology/>.
48. Augur, Hannah. Dataconomy. *Beginner's Guide to the History of Data Science*. [En línea] 11 de Marzo de 2016. <https://dataconomy.com/2016/03/beginners-guide-history-data-science/>.
49. 9 awesome difference between Data Science and Data mining. *Educba*. [En línea] [Citado el: 15 de Febrero de 2020.] <https://www.educba.com/data-science-vs-data-mining/>.
50. Gour, Rinu. 8 key differences between data science and data mining. [En línea] 2019. [Citado el: 15 de Febrero de 2020.] <https://becominghuman.ai/8-key-differences-between-data-science-and-data-mining-674f09599df2>.
51. PowerData. *Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad*. [En línea] [Citado el: 15 de Febrero de 2020.] <https://www.powerdata.es/big-data>.

52. Ladrero, Iñaki. Baoss Analytics Everywhere. *10 Ejemplos de usos reales de Big Data Analytics*. [En línea] 19 de abril de 2017. <https://www.baoss.es/10-ejemplos-usos-reales-big-data/>.
53. MaxBrainFunction . *What is Strong AI and what are some examples?* [En línea] 2018. [Citado el: 10 de Febrero de 2020.] <https://maxbrainfunction.com/strong-ai/>.
54. Wu, Jun. AI, Machine Learning, Deep Learning explained simply. [En línea] 2019. [Citado el: 10 de Febrero de 2020.] <https://towardsdatascience.com/ai-machine-learning-deep-learning-explained-simply-7b553da5b960>.
55. Rouse, Margaret. Artificial Intelligence. [En línea] 2019. [Citado el: 10 de Febrero de 2020.] <https://searchenterpriseai.techtarget.com/definition/AI-Artificial-Intelligence>.
56. Frankenfield, Jake. Artificial Intelligence. [En línea] Enero de 2020. [Citado el: 10 de Febrero de 2020.] <https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp>.
57. Jeffcock, Pete. Demystifying machine learning: an overview. [En línea] 2017. [Citado el: 12 de Febrero de 2020.] <https://blogs.oracle.com/bigdata/demystifying-machine-learning%3a-an-overview>.
58. Brownlee, Jason. What is deep learning? [En línea] 2019. [Citado el: 12 de Febrero de 2020.] <https://machinelearningmastery.com/what-is-deep-learning/>.
59. Hargrave, Marshall. Deep learning. [En línea] 2019. [Citado el: 16 de Febrero de 2020.] <https://www.investopedia.com/terms/d/deep-learning.asp>.
60. Chollet, Francois. *Deep Learning with Python*. 2018.
61. Ligdi Gonzalez. *Regresión Lineal Múltiple – Teoría*. [En línea] 21 de Diciembre de 2018. [Citado el: 29 de Enero de 2020.] <https://ligdigonzalez.com/regresion-lineal-multiple-machine-learning-teoria/>.
62. Ligdi Gonzales. *Regresión Polinomial – Teoría*. [En línea] 11 de Enero de 2019. [Citado el: 10 de Enero de 2020.] <https://ligdigonzalez.com/algoritmo-regresion-polinomial-machine-learning/>.
63. Towards Data Science. *Machine Learning Basics with the K-Nearest Neighbors Algorithm*. [En línea] 10 de Septiembre de 2018.

- <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>.
64. K Nearest Neighbors - Classification. [En línea] [Citado el: 17 de Enero de 2020.] https://www.saedsayad.com/k_nearest_neighbors.htm.
65. Aprende Machine Learning. *Arbol de Decisión en Python: Clasificación y predicción*. [En línea] 3 de Abril de 2018. [Citado el: 12 de Enero de 2020.] <https://www.aprendemachinelearning.com/arbol-de-decision-en-python-clasificacion-y-prediccion/>.
66. IBM knowledge Center. El modelo de redes neuronales. [En línea] [Citado el: 11 de Febrero de 2020.] https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/components/neuralnet/neuralnet_model.html.
67. Olivera, Oscar Garcia-Olalla. Redes neuronales artificiales: Que son y como se entrenan Parte I. [En línea] 11 de Febrero de 2020. <https://www.xeridia.com/blog/redes-neuronales-artificiales-que-son-y-como-se-entrenan-parte-i>.
68. Innovation., Atria. Que son las redes neuronales y sus funciones. [En línea] [Citado el: 11 de Febrero de 2020.] <https://www.atriainnovation.com/que-son-las-redes-neuronales-y-sus-funciones/>.
69. Julián, Guillermo. Las redes neuronales: qué son y por qué están volviendo. [En línea] [Citado el: 11 de Febrero de 2020.] <https://www.xataka.com/robotica-e-ia/las-redes-neuronales-que-son-y-por-que-estan-volviendo> .
70. Dendrita. *Wikipedia*. [En línea] [Citado el: 8 de Febreo de 2020.] <https://es.wikipedia.org/wiki/Dendrita>.
71. Axón. *Wikipedia*. [En línea] [Citado el: 8 de Febrero de 2020.] <https://es.wikipedia.org/wiki/Ax%C3%B3n>.
72. Neurona de McCulloch-Pitts. *Wikipedia*. [En línea] [Citado el: 11 de Febrero de 2020.] https://es.wikipedia.org/wiki/Neurona_de_McCulloch-Pitts .

73. Coursera. Artificial neural networks. *Introduction to Deep learning with keras*. [En línea] [Citado el: 11 de Febrero de 2020.] <https://www.coursera.org/learn/introduction-to-deep-learning-with-keras/lecture/HOKuA/artificial-neural-networks>.
74. —. Backpropagation. *Introduction to deep learning with keras*. [En línea] [Citado el: 10 de Febrero de 2020.] <https://www.coursera.org/learn/introduction-to-deep-learning-with-keras/lecture/lfbxX/backpropagation>.
75. —. Activation functions. *Introduction to deep learning with keras*. [En línea] [Citado el: 12 de Febrero de 2020.] <https://www.coursera.org/learn/introduction-to-deep-learning-with-keras/lecture/tL4uI/activation-functions>.
76. —. Gradient Descent. *Introduction to deep learning with keras*. [En línea] [Citado el: 8 de Febrero de 2020.] <https://www.coursera.org/learn/introduction-to-deep-learning-with-keras/lecture/xabij/gradient-descent>.
77. Wang, Chi-Feng. Finding the cost function of neural networks. *Towards Data Science*. [En línea] 2018. [Citado el: 8 de Febrero de 2020.] <https://towardsdatascience.com/step-by-step-the-math-behind-neural-networks-490dc1f3cfd9>.
78. —. The beginner's guide to gradient descent. *Towards Data science*. [En línea] 2018. [Citado el: 8 de Febrero de 2020.] <https://towardsdatascience.com/the-beginners-guide-to-gradient-descent-c23534f808fd>.
79. S., Edgar A. Dávila. *FACTIBILIDAD TÉCNICO-ECONÓMICA DE LA APLICACIÓN DE LOS DISPOSITIVOS DE CONTROL DE INFLUJO AUTÓNOMOS Y EL IMPACTO DEL ARENAMIENTO EN SU DESEMPEÑO A SER EVALUADAS EN LAS ARENAS "D" DE LA FORMACION OFICINA, BLOQUE JUNIN*. Caracas : s.n., 2015.
80. Eliana Alvarado, Edgar Dávila. *Producción de Arena*. [Presentación PowerPoint] Caracas : s.n., 2015.
81. Manejo de petróleo y gas en superficie. *Mecanismos de empuje natural*. [En línea] 23 de diciembre de 2013. <http://manejodepetroleoygas.blogspot.com/2013/12/mecanismos-de-empuje-natural.html>.

82. Espinosa, Juan Manuel Gonzales. Daño a la formación en pozos petroleros. México D.F. : s.n., 2014, pág. 114.
83. Terminación y mantenimiento de pozos. *Oil Production*. [En línea] http://oilproduction.net/files/manual_terminacion_de_pozos.pdf.
84. Curvas de declinación de producción. *La comunidad petrolera*. [En línea] 23 de Junio de 2009. <https://www.lacomunidadpetrolera.com/2009/06/curvas-de-declinacion-de-produccion.html>.
85. Noticias, PDVSA. Venezuela posee las reservas de crudos más grandes del mundo. *PDVSA Oficial*. [En línea] 28 de Marzo de 2006. http://www.pdvsa.com/index.php?option=com_content&view=article&id=1501:2423&catid=10&Itemid=589&lang=es.
86. Silva, Angel Da. La Comunidad petrolera. *Mecanismos de producción*. [En línea] 22 de marzo de 2009. <https://www.lacomunidadpetrolera.com/2009/03/mecanismos-de-empuje-natural-en.html>.
87. Romero, Jorge. Jorge Romero. *Metodología de la Ciencia de Datos*. [En línea] 11 de Junio de 2019. [Citado el: 2 de Enero de 2020.] <https://jorgeromero.net/metodologias-de-mineria-de-datos/>.
88. Olivera, Oscar Garcia-Olalla. Redes neuronales artificiales: Que son y como se entrenan. [En línea] [Citado el: 11 de Febrero de 2020.] <https://www.xeridia.com/blog/redes-neuronales-artificiales-que-son-y-como-se-entrenan-parte-i>.
89. Carlos Curtis, et. all. Yacimientos de petróleo pesado. *Oilproduction.net/*. [En línea] [Citado el: 15 de marzo de 2020.] http://www.oilproduction.net/files/petroleos_pesados.pdf.