

UNIVERSIDAD MAYOR DE SAN ANDRÉS
FACULTAD DE INGENIERÍA
INGENIERÍA PETROLERA



TRABAJO DIRIGIDO

**APLICACIÓN DEL MÉTODO MACHINE
LEARNING EN LA TOMA DE DECISIONES
PARA OPTIMIZAR LA PERFORACIÓN DE
POZOS**

Univ. Luis Erwin Condori Mamani

Tutor: Ing. Wilson Medinaceli Mendoza

**Asesor Institucional: Ing. Mauricio Jesús Ortuño
Riveros**

**Identidad Institucional: Agencia Nacional de
Hidrocarburos A.N.H.**

La Paz, Bolivia

Año 2024



UNIVERSIDAD MAYOR DE SAN ANDRÉS
FACULTAD DE INGENIERÍA



La Facultad de Ingeniería de la Universidad Mayor de San Andrés autoriza el uso de la información contenida en este documento si los propósitos son estrictamente académicos.

LICENCIA DE USO

El usuario está autorizado a:

- Visualizar el documento mediante el uso de un ordenador o dispositivo móvil.
- Copiar, almacenar o imprimir si ha de ser de uso exclusivamente personal y privado.
- Copiar textualmente fragmentos del texto para su uso en trabajos académicos (citando debidamente la fuente).

AGENCIA NACIONAL DE HIDROCARBUROS
DIRECCIÓN DE OBSERVACIÓN ENERGÉTICA
REGULATORIA



TRABAJO DIRIGIDO

**APLICACIÓN DEL MÉTODO MACHINE
LEARNING EN LA TOMA DE DECISIONES
PARA OPTIMIZAR LA PERFORACIÓN DE
POZOS**

Univ. Luis Erwin Condori Mamani

Tutor: Ing. Wilson Medinaceli Mendoza

**Asesor Institucional: Ing. Mauricio Jesús Ortuño
Riveros**

La Paz, Bolivia

Año 2024

DEDICATORIA

A mis queridos padres, Valentina y Damian,
por su amor incondicional, su sacrificio y su fe inquebrantable
en mí.

Su ejemplo de vida ha sido mi guía y mi inspiración constante.

A mi hermana,

por su apoyo y comprensión,

y por ser la luz que ilumina cada paso de mi camino.

AGRADECIMIENTO

En este camino que culmina con la finalización de mi trabajo dirigido, deseo expresar mi gratitud a todas las personas e instituciones que han contribuido a mi formación y al éxito de este proyecto.

Mi sincero agradecimiento al Ing. Wilson Medinaceli Mendoza, tutor de carrera, por su invaluable asesoramiento, paciencia y dedicación, que han sido clave en mi proceso de aprendizaje y en la consecución de mis metas académicas.

Igualmente, agradezco al Ing. Mauricio Jesús Ortuño Riveros, tutor institucional en la Agencia Nacional de Hidrocarburos (ANH), cuya guía experta y apoyo constante han sido esenciales para la realización de este trabajo dirigido.

No puedo dejar de mencionar el apoyo incondicional de mi familia y amigos, quienes con su amor, comprensión y aliento, me han acompañado en cada paso de este viaje.

Extiendo un especial reconocimiento a la Universidad Mayor de San Andrés por ser el pilar fundamental de mi educación superior y por brindarme las herramientas y conocimientos necesarios para mi desarrollo profesional.

Finalmente, mi gratitud a la Agencia Nacional de Hidrocarburos por permitirme desarrollar mi trabajo en un entorno desafiante y enriquecedor, contribuyendo significativamente a mi crecimiento personal y profesional.

Índice general

| | |
|---|-------------|
| Índice de Figuras | VI |
| Índice de Tablas | VIII |
| Lista de Términos | X |
| Resumen Ejecutivo | 1 |
| 1. GENERALIDADES | 3 |
| 1.1. Introducción | 3 |
| 1.2. Antecedentes | 4 |
| 1.3. Marco Normativo | 5 |
| 1.4. Planteamiento del Problema | 7 |
| 1.4.1. Identificación del Problema | 7 |
| 1.4.2. Formulación del Problema | 8 |
| 1.5. Objetivos y Acciones | 9 |
| 1.5.1. Objetivo General | 9 |
| 1.5.2. Objetivos Específicos y Acciones de la Investigación | 9 |
| 1.6. Justificación | 9 |
| 1.6.1. Justificación Técnica | 9 |
| 1.6.2. Justificación Económica | 10 |

| | |
|---|-----------|
| 1.6.3. Justificación Ambiental | 10 |
| 1.7. Alcance y Límites | 11 |
| 1.7.1. Alcances | 11 |
| 1.7.2. Límitaciones | 11 |
| 2. MARCO TEÓRICO | 13 |
| 2.1. Conceptos Básicos de la Perforación de Pozos | 13 |
| 2.2. Tipos de Pozos | 15 |
| 2.2.1. Pozos de Investigación Estratigráfica | 15 |
| 2.2.2. Pozos Exploratorios | 15 |
| 2.2.3. Pozos de Desarrollo | 16 |
| 2.2.4. Pozos de Avanzada | 16 |
| 2.3. Perforación de Pozos | 16 |
| 2.3.1. Coordenadas del Conductor y Objetivos | 17 |
| 2.3.2. Posición Estructural | 17 |
| 2.3.3. Sarta de Perforación | 17 |
| 2.3.4. Parámetros de Perforación | 18 |
| 2.4. Fluidos de Perforación | 26 |
| 2.4.1. Funciones del Fluido de Perforación | 26 |
| 2.5. Balance de un Pozo | 27 |
| 2.5.1. Perforación Bajo Balance | 28 |
| 2.5.2. Perforación Sobre Balance | 28 |
| 2.6. Control de Pozos | 29 |
| 2.6.1. Presión | 29 |
| 2.6.2. Presión de Fluido | 30 |
| 2.6.3. Presion Hidrostática | 30 |
| 2.6.4. Presión de Formación | 31 |
| 2.6.5. Presión de Fractura | 32 |

| | | |
|-----------|--|-----------|
| 2.6.6. | Presión de Fondo de Pozo | 33 |
| 2.6.7. | Densidad Equivalente de Circulación | 34 |
| 2.7. | Parámetros de Perforación e Influencia en la ROP | 35 |
| 2.8. | Modelos de ROP | 35 |
| 2.8.1. | Modelos Mecánicos | 35 |
| 2.8.2. | Modelos Empíricos | 38 |
| 2.8.3. | Modelos de Regresión Lineal | 41 |
| 2.8.4. | Modelos de Redes Neuronales | 43 |
| 2.9. | Exponente D de Perforación | 46 |
| 3. | ALGORITMOS DE MACHINE LEARNING PARA LA OPTIMIZACIÓN DE LA | |
| | ROP | 49 |
| 3.1. | Machine Learning | 49 |
| 3.1.1. | Machine Learning Supervisado | 52 |
| 3.1.2. | Machine Learning No Supervisado | 52 |
| 3.1.3. | Machine Learning por Refuerzo | 53 |
| 3.1.4. | IA Generativa | 55 |
| 3.2. | Herramientas y Librerías de Machine Learning | 56 |
| 3.2.1. | Python | 56 |
| 3.2.2. | Librerías para Ciencia de Datos | 56 |
| 3.2.3. | Librerías para la Visualización de Datos | 58 |
| 3.2.4. | Librerías para Machine Learning | 59 |
| 3.3. | Análisis Exploratorio de Datos | 59 |
| 3.4. | Algoritmos Machine Learning | 61 |
| 3.4.1. | Algoritmo Decision Tree | 62 |
| 3.4.2. | Algoritmo Gradient Boosting Machine | 64 |
| 3.4.3. | Algoritmo eXtreme Gradient Boosting | 66 |
| 3.4.4. | Algoritmo Random Forest | 67 |

| | |
|---|------------|
| 3.5. Métricas de Evaluación de un Modelo ML | 69 |
| 3.5.1. Error Cuadrático Medio | 70 |
| 3.5.2. Error Absoluto Medio | 70 |
| 3.5.3. Raíz del Error Cuadrático Medio | 70 |
| 3.5.4. Coeficiente de Determinación R2 | 70 |
| 4. MARCO APLICATIVO | 72 |
| 4.1. Introducción | 72 |
| 4.2. Descripción Geológica | 74 |
| 4.2.1. Campo San Antonio | 74 |
| 4.2.2. Campo Santa Rosa | 76 |
| 4.2.3. Campo Rio Grande | 77 |
| 4.3. Análisis Exploratorios de Datos | 78 |
| 4.3.1. Análisis de Datos | 78 |
| 4.4. Modelo predictivo Machine Learning | 86 |
| 4.4.1. Modelo predictivo con el algoritmo GBM | 86 |
| 4.4.2. Modelo predictivo con el algoritmo xGBoost | 88 |
| 4.4.3. Modelo Predictivo con el Algoritmo Random Forest | 89 |
| 4.5. Aplicación de los Modelos en el Pozo RGD-100D | 92 |
| 5. ANÁLISIS DE COSTOS | 95 |
| 5.1. Estimación de Costo | 95 |
| 5.2. Ventajas y Desventajas | 96 |
| 5.3. Tiempo No Productivo NPT | 98 |
| 5.4. Análisis de Costos | 98 |
| 6. CONCLUSIONES Y RECOMENDACIONES | 101 |
| 6.1. Conclusiones | 101 |
| 6.2. Recomendaciones | 103 |

| | |
|--|------------|
| Bibliografía | 104 |
| A. Fundamentos de Estadística y Análisis de Datos con Python | 108 |
| A.1. Introducción a la Estadística | 108 |
| A.1.1. Definición y Importancia | 108 |
| A.1.2. Tipos de Datos Estadísticos | 109 |
| A.1.3. Fórmulas Básicas en Estadística | 109 |
| A.2. Análisis Exploratorio de Datos | 110 |
| A.2.1. Objetivos del EDA | 110 |
| A.2.2. Técnicas de EDA | 110 |
| A.2.3. Herramientas de Visualización | 111 |
| A.2.4. Creación de Mapas de Calor y Gráficas de Dispersión | 113 |
| A.2.5. Estadísticas Descriptivas | 114 |
| A.2.6. Detección de Valores Atípicos | 114 |
| A.2.7. Correlación y Causalidad | 114 |
| A.3. Fundamentos de Machine Learning | 115 |
| A.3.1. Tipos de Aprendizaje en Machine Learning | 115 |
| A.3.2. Modelos y Algoritmos Comunes | 115 |
| A.3.3. Evaluación de Modelos | 116 |
| A.3.4. Procesamiento de Lenguaje Natural | 116 |
| A.4. Aplicaciones Prácticas con Python | 117 |
| A.4.1. Bibliotecas de Python para Estadística y ML | 117 |
| A.4.2. Casos de Estudio y Ejemplos de Código | 118 |
| A.4.3. Mejores Prácticas y Consideraciones Éticas | 119 |

Índice de figuras

| | |
|---|----|
| 2.1. Sistema de Rotación | 14 |
| 2.2. Presion Diferencial | 29 |
| 2.3. Relación de la ROP de acuerdo al Modelo de Bourgoyne y Young | 41 |
| 2.4. Rate of Penetration (ROP) Vs Revoluciones por minuto (RPM) | 46 |
| 2.5. Rate of Penetration (ROP) Vs Peso sobre el trépano (WOB) | 46 |
| 2.6. Exponente D y Dc | 48 |
| 2.7. Comparación entre el exponente D y Dc | 48 |
| 3.1. Aplicaciones del ML | 51 |
| 3.2. Grupos de clústeres con demarcaciones naturales | 53 |
| 3.3. Proceso de Desarrollo del EDA | 60 |
| 4.1. Desarrollo del modelo ML para la optimización de la ROP | 73 |
| 4.2. Campo San Antonio | 75 |
| 4.3. Campo Santa Rosa | 76 |
| 4.4. Campo Rio Grande | 78 |
| 4.5. Parámetros de Perforación de Pozos | 80 |
| 4.6. Histograma de la Rate of Penetration (ROP) | 81 |
| 4.7. Diagrama de dispersión de los parámetros de perforación | 82 |
| 4.8. Correlación de los Parámetros de Perforación | 84 |
| 4.9. Diagrama de cajas de los Parámetros de Perforación | 85 |

| | |
|--|-----|
| 4.10. Modelo Gradient Boosting Machine (GBM) | 87 |
| 4.11. Modelo eXtreme Gradient Boosting (xGBoost) | 89 |
| 4.12. Modelo Random Forest | 91 |
| 4.13. Gráfica de la ROP del Pozo RGD-100D | 93 |
| 4.14. Gráfica del Exponente Dc del Pozo RGD-100D | 94 |
| 5.1. Curva de Avance en el Campo Sábalo | 99 |
| 5.2. Distribución de Tiempos de Perforación | 100 |
| 5.3. Distribución de NPT | 100 |
| A.1. Gráficos de barras y histogramas | 111 |
| A.2. Diagramas de Cajas | 111 |
| A.3. Gráficos de dispersión | 112 |
| A.4. Gráficos de líneas | 112 |
| A.5. Mapa de calor | 113 |

Índice de tablas

| | |
|---|----|
| 2.1. Modelos Analíticos de la ROP | 45 |
| 3.1. Ventajas y Desventajas del Algoritmo DT | 63 |
| 3.2. Ventajas y Desventajas del Algoritmo GBM | 65 |
| 3.3. Ventajas y Desventajas del Algoritmo xGBoost | 67 |
| 3.4. Ventajas y Desventajas del Algoritmo RF | 69 |
| 4.1. Registro de Parámetros de Perforación | 80 |
| 4.2. Tabla de Comparación de los Modelos ML | 92 |
| 4.3. Tabla de Métricas de los Modelos ML | 92 |
| 5.1. Estimación de Costos | 96 |
| 5.2. Costo de Perforación en el Campo Sábalo | 99 |

Acronimos

AI Artificial Intelligence.

ANH Agencia Nacional de Hidrocarburos.

ANN Red Neuronal Artificial.

BHA Bottom Hole Assembly.

BHP Bottom Hole Pressure.

BOP Blowout Preventer.

CPE Constitución Política del Estado.

DC Drill Collar.

DS Decreto Supremo.

DT Decision Tree.

ECD Densidad Equivalente de Circulacion.

EDA Exploratory Data Analysis.

GBM Gradient Boosting Machine.

LWD Logging While Drilling.

MAE Mean Absolute Error.

ML Machine Learning.

MSE Mean Squared Error.

MW Mud Weight.

MWD Measuring While Drilling.

NLP Natural Language Processing.

NPT Non-Productive Time.

PCA Análisis de Componentes Principales.

PDC Diamond Compact Bits.

PDF Presion de Fractura.

PDFr Presion de Fricción.

PFm Presion de Formacion.

PH Presion Hidrostatica.

pH Grado de alcalinidad de una sustancia o una solución.

R2 Coeficiente de Determinación.

RF Random Forest.

RMSE Root Mean Squared Error.

ROC Curve Receiver Operating Characteristic Curve.

ROP Rate of Penetration.

RPM Revoluciones por minuto.

TVD Profundidad Vertical Verdadera.

WOB Peso sobre el trépano.

xGBoost eXtreme Gradient Boosting.

YPFB Yacimientos Petrolíferos Fiscales Bolivianos.

Resumen Ejecutivo

Este trabajo detalla el desarrollo de un modelo Machine Learning (ML) para la optimización de la Rate of Penetration (ROP) (Tasa de Penetración) en la perforación de pozos petroleros. Este trabajo se centra en la implementación y comparación de algoritmos avanzados de Machine Learning (ML), específicamente Gradient Boosting Machine (GBM), eXtreme Gradient Boosting (xGBoost) y Random Forest (RF), con el objetivo de mejorar la eficiencia operativa y reducir costos. La investigación se basa en datos operativos recopilados de cuatro pozos distintos: RGD-53, RGD-100D, SBL-8 y SRW-8D, proporcionando una base sólida para el entrenamiento y la validación de los modelos propuestos.

Los resultados obtenidos demuestran que los modelos ML son herramientas poderosas para predecir y optimizar la ROP. El algoritmo xGBoost, en particular, mostró un rendimiento superior, seguido por RF y GBM, indicando su potencial para ser implementado en operaciones de perforación reales. Estos hallazgos sugieren que la integración de modelos predictivos basados en ML puede resultar en una planificación más precisa y una ejecución más eficiente de la perforación de pozos petroleros.

Las conclusiones del estudio enfatizan la importancia de la selección de algoritmos y la calidad de los datos en la precisión de las predicciones de la Rate of Penetration (ROP). La aplicación práctica de estos modelos podría significar un avance considerable en la industria petrolera, permitiendo una toma de decisiones más informada y

precisa, así como también una optimización efectiva de los recursos. Además, se recomienda la realización de pruebas adicionales en un entorno de producción para validar la eficacia de los modelos en tiempo real y explorar su integración con sistemas de control automatizado.

Este trabajo no solo contribuye al campo académico proporcionando una metodología robusta para la optimización de la ROP, sino que también ofrece una perspectiva aplicable para la mejora continua en las operaciones de perforación. La implementación de estos modelos tiene el potencial de transformar las prácticas actuales, llevando a la industria hacia una era de mayor automatización y eficiencia.

Capítulo 1

GENERALIDADES

1.1. Introducción

La industria petrolera en Bolivia ha sido un pilar fundamental en la economía del país, con campos como Río Grande, San Antonio y Santa Rosa, Bolivia ha establecido una sólida infraestructura de extracción y procesamiento de hidrocarburos. Este trabajo se centra en la aplicación de técnicas avanzadas de *Machine Learning (ML)* para optimizar la Tasa de Penetración, del inglés *Rate of Penetration (ROP)*, un indicador crítico de eficiencia en la perforación de pozos petroleros.

A pesar de los avances tecnológicos, la optimización de la ROP sigue siendo un desafío debido a la complejidad geológica y a las variables operativas involucradas. Este trabajo aborda la necesidad de mejorar la precisión en la estimación de la ROP para maximizar la eficiencia y reducir los costos operativos en los pozos RGD-100D, RGD-53, SBL-8 y SRW-8, por lo que el objetivo principal es desarrollar un modelo predictivo de *Machine Learning* que mejore la estimación de la ROP en los pozos seleccionados. Los objetivos secundarios incluyen analizar las variables que influyen durante la perforación y las correlaciones entre las variables de los parámetros de

perforación de pozos con la ROP y evaluar la aplicabilidad del modelo en diferentes escenarios de perforación.

La relevancia de este trabajo radica en su potencial para transformar las operaciones de perforación en Bolivia, ofreciendo una herramienta para la toma de decisiones basada en datos y mejorando la competitividad de la industria petrolera boliviana en el mercado global, por lo que la aplicación de algoritmos de ML puede resultar en una estimación más precisa de la ROP en comparación con los métodos tradicionales. Entonces se utilizarán técnicas de ML, como Gradient Boosting Machine (GBM), eXtreme Gradient Boosting (xGBoost) y Random Forest (RF), para analizar los datos históricos de los pozos y desarrollar el modelo predictivo. La metodología incluirá la limpieza de datos, la selección de características y la validación cruzada del modelo. El trabajo se centrará en los pozos petroleros RGD-100D, RGD-53, SBL-8 y SRW-8 en Bolivia.

1.2. Antecedentes

En los últimos años el avance de la tecnología de la perforación petrolera ha suscitado un incremento sustancial en operaciones dentro de la industria petrolera, y más aun con la implementación de nuevas tecnologías computacionales como la Inteligencia Artificial, *del inglés Artificial Intelligence (AI)*.

La exploración y explotación de hidrocarburos en Bolivia ha sido una actividad de gran relevancia económica y social desde el siglo XX. Los campos petrolíferos de Río Grande, San Antonio y Santa Rosa han jugado un papel crucial en el desarrollo energético del país. Estos campos, descubiertos y desarrollados a lo largo de las últimas décadas, han contribuido significativamente a la economía nacional, alternando entre la producción de petróleo y gas natural.

En el contexto de la industria petrolera, la optimización de la ROP es un aspecto técnico esencial que impacta directamente en la eficiencia y costos de perforación. Tradicionalmente, la ROP se ha optimizado mediante métodos empíricos y la experiencia de los ingenieros en campo. Sin embargo, con el advenimiento de la era digital y el auge de la AI, nuevas metodologías están emergiendo para enfrentar este desafío.

El uso de ML en la optimización de la ROP representa una innovación disruptiva en la industria petrolera. A través del análisis de grandes volúmenes de datos históricos y la aplicación de modelos predictivos, es posible mejorar la precisión en la estimación de la ROP, lo que conduce a operaciones más eficientes y rentables.

En Bolivia, la aplicación de AI en la industria petrolera aún está en sus etapas iniciales. Sin embargo, proyectos como el de Microsoft en colaboración con YPFB Chaco demuestran el potencial de estas tecnologías para transformar el sector. La implementación de soluciones basadas en AI para la optimización de procesos no solo es una tendencia global, sino también una necesidad local para mantener la competitividad en un mercado cada vez más tecnológico (Microsoft, 2024).

Este trabajo se inscribe en este contexto innovador, buscando aplicar técnicas de ML para la optimización de la ROP en los pozos petroleros de Bolivia. Los pozos RGD-100D, RGD-53, SBL-8 y SRW-8, seleccionados para este estudio, representan una muestra diversa de la infraestructura petrolera boliviana y ofrecen una oportunidad única para demostrar la aplicabilidad y beneficios de la AI en la optimización de la perforación.

1.3. Marco Normativo

Constitución Política del Estado (CPE), en su Artículo 365 dispone que: *“una institución autárquica de derecho público con autonomía de gestión administrativa,*

técnica y económica, bajo la tuición del ministerio del ramo, será responsable de regular, controlar, supervisar y fiscalizar las actividades de toda la cadena productiva hasta la industrialización, en el marco de la política estatal de hidrocarburos conforme con la ley”.

Ley N° 466 de Empresa Pública del 26 de diciembre de 2013, que establece en su Disposición Final Séptima que: *“para el cumplimiento de lo dispuesto en el Artículo 365 de la Constitución Política del Estado, la Agencia Nacional de Hidrocarburos (ANH) queda encargada de emitir la normativa técnico jurídica necesaria para el cumplimiento de sus atribuciones de regulación, control, supervisión y fiscalización de todas las actividades del circuito productivo”.*

Ley N° 767 de *“Promoción para la Inversión en Exploración y Explotación Hidrocarburífera”* del 11 de diciembre de 2015, que tiene por objeto promover las inversiones en las actividades de exploración y explotación de hidrocarburos, mismas que se declaran de interés nacional en todo el territorio del Estado Plurinacional de Bolivia.

Decreto Supremo (DS) N° 4616, del 10 de noviembre de 2021, que tiene por objeto *“(. . .) establecer condiciones para la aplicación de incentivos a las inversiones en Exploración y Explotación Hidrocarburífera cuando Yacimientos Petrolíferos Fiscales Bolivianos (YPFB) sea operador (. . .)”*, El Parágrafo V del Artículo 2 (MODIFICACIONES) del D.S. N° 4616, que cuando YPFB opere por sí misma, la ANH cumplirá la función de: **“6. Fiscalizar los volúmenes producidos sujetos a incentivo”**, modificando el Decreto Supremo (DS) N° 2830 Párrafo II del Artículo 15.

Resolución Ministerial (R.M.) N° 0183/2021 de fecha 30 de diciembre de 2021, donde el Ministerio de Hidrocarburos y Energías aprueba el **“REGLAMENTO PARA ESTABLECER LOS PROCEDIMIENTOS PARA LA APLICACIÓN DEL INCENTIVO A TITULARES DE LOS CONTRATOS DE SERVICIOS PETROLEROS Y YACIMIENTOS PETROLÍFEROS FISCALES BOLIVIANOS CUANDO OPERE POR SÍ MISMA”**

de acuerdo al **Artículo 30 (FISCALIZACIÓN DE ACTIVIDADES EN CAMPOS)** se establece:

1. *Cuando YPFB opere por sí misma, la ANH podrá designar personal técnico in situ para fiscalizar la producción de las áreas operadas por YPFB en cumplimiento de las atribuciones otorgadas en el Artículo 15 del Decreto Supremo N° 2830, modificado por el Parágrafo V del Artículo 2 del Decreto Supremo N° 4616.*
2. *YPFB ofrecerá las condiciones necesarias en Campo, para que el personal de la ANH cumpla las funciones de fiscalización de los volúmenes sujetos a incentivos.*

1.4. Planteamiento del Problema

1.4.1. Identificación del Problema

La creciente demanda global y la continua dependencia de los hidrocarburos han impulsado a la industria petrolera a buscar innovaciones en la exploración y explotación de estos recursos. En la actualidad, la búsqueda de petróleo se extiende a profundidades que superan con creces las alcanzadas hace tres décadas, lo que presenta nuevos desafíos técnicos y operativos (Short, 1993).

En este contexto, la optimización de la ROP durante la perforación de pozos petroleros emerge como un aspecto crítico que impacta directamente en la eficiencia y los costos asociados al proceso de extracción. Históricamente, la ROP se ha mejorado a través de métodos empíricos basados en la experiencia acumulada de los ingenieros de perforación. No obstante, la precisión y eficiencia de estos enfoques tradicionales son a menudo cuestionables, lo que puede conducir a operaciones de perforación extendidas y un incremento en los costos operacionales.

El presente trabajo se enfoca en la aplicación de técnicas de Machine Learning (ML) para predecir y optimizar la ROP, con un estudio aplicado en los pozos SBL-8, RGD-53, RGD100D y SRW-8D, todos situados en Bolivia. La meta es superar las limitaciones de los métodos convencionales y proporcionar una herramienta analítica que permita una toma de decisiones más informada y eficiente en la perforación de pozos petroleros.

1.4.2. Formulación del Problema

La optimización de la ROP en la perforación de pozos petroleros es un desafío multifacético que afecta tanto la eficiencia operativa como los costos de producción. En el escenario actual, donde la perforación se lleva a cabo en condiciones cada vez más complejas y a mayores profundidades, se requiere una metodología que permita predecir la ROP con mayor precisión y rapidez.

El problema central que aborda este trabajo es: ¿Cómo podemos mejorar la predicción de la ROP utilizando técnicas de Machine Learning para reducir los tiempos de perforación? Este problema se investigará en el contexto de los pozos SBL-8, RGD-53, RGD-100D y SRW-8D en Bolivia, donde se recopilarán y analizarán datos históricos y en tiempo real para desarrollar un modelo predictivo robusto.

La formulación del problema incluye la identificación de variables clave que influyen en la ROP, la selección de algoritmos de machine learning adecuados y la evaluación de la precisión y eficacia del modelo propuesto. El objetivo es proporcionar una herramienta analítica que mejore la toma de decisiones en la perforación de pozos y que pueda adaptarse a diferentes entornos geológicos y operativos.

1.5. Objetivos y Acciones

1.5.1. Objetivo General

Desarrollar un modelo avanzado de Machine Learning que, al ser aplicado en la perforación de pozos de Bolivia, mejore significativamente la toma de decisiones operativas.

1.5.2. Objetivos Específicos y Acciones de la Investigación

- Comprender, analizar y valorar las variables que afectan a la eficiencia en el avance de perforación, para interpretar su impacto y proponer medidas de optimización.
- Investigar la tecnología del Machine Learning, así como también la implementación en perforación de pozos para identificar prácticas óptimas de implementación.
- Desarrollar un modelo predictivo de Machine Learning, validado con un Conjunto de Datos (Datasets) representativos que pueda predecir con precisión el avance de la perforación y facilitar la toma de decisiones en tiempo real.

1.6. Justificación

1.6.1. Justificación Técnica

A medida que las exploraciones hidrocarburíferas se las hacen a mayores profundidades, el riesgo técnico y económico también aumentan considerablemente. Un modelo ML no solo agiliza el tratamiento de los datos adquiridos, sino también en la optimización del avance de perforación de un pozo que además genera una mayor confiabilidad en la toma de decisiones ya sea para una mejor optimización del tiempo o bien

ante situaciones imprevistas.

1.6.2. Justificación Económica

El alcance de la investigación se centra en la aplicación y validación de los modelos GBM, xGBoost y Random Forest para optimizar la toma de decisiones en la perforación de pozos. Aunque el trabajo no incluye un análisis económico directo, la justificación económica se deriva de la potencial eficiencia operativa y mejora en la toma de decisiones que el modelo puede proporcionar. A largo plazo, estas mejoras técnicas podrían traducirse en una reducción de costos y un aumento en la rentabilidad de las operaciones de perforación.

1.6.3. Justificación Ambiental

La aplicación del ML en la perforación de pozos ofrece una oportunidad para optimizar la eficiencia y por ende minimizar el impacto ambiental. Al predecir con precisión la ROP, se evita la perforación excesiva y el consumo innecesario de recursos por lo que cumplir con el Reglamento Ambiental para el Sector Hidrocarburos (DS 24335) es fundamental para garantizar operaciones sostenibles y responsables, es más, la optimización de la perforación mediante el uso de ML no solo ofrece beneficios económicos, sino que también puede contribuir a la protección ambiental.

1.7. Alcance y Límites

1.7.1. Alcances

Alcance Temático

El área de aplicación que alcanza el presente trabajo, abarca la explotación de hidrocarburos en la industria petrolera, más específicamente en el área de perforación de pozos de petróleo y/o gas en Bolivia.

Alcance Operativo

Alcance de Investigación

El trabajo está enfocado a la validación técnica y metodológica del Modelo ML, así como su desarrollo, configuración y análisis del rendimiento en base a datos históricos de pozos de Bolivia y evaluación de la precisión del modelo en la predicción de resultados. Por otra parte, su aplicación práctica del modelo en la optimización de la perforación de pozos dependerá de factores de diseño de software (Persistencia del Modelo) ya que su implementación en un software comercial requiere desde la evaluación de la arquitectura de software hasta el estudio de mercado, algo que está fuera del alcance del trabajo, por lo que esto incluye el análisis económico.

1.7.2. Límites

Limitaciones Metodológicas

El modelo Gradient Boosting Machine (GBM) es un algoritmo muy eficaz a grandes volúmenes de datos, pero tiene limitaciones, como la tendencia al sobreajuste si no se configura adecuadamente. Además, la interpretación de los modelos GBM puede ser compleja, lo que dificulta la explicación de las decisiones del modelo a

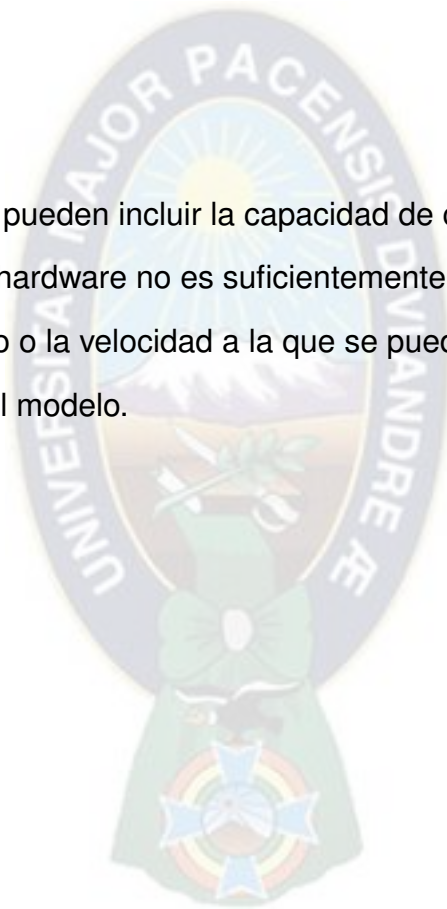
personas no técnicas.

Limitaciones de Datos

Una limitación común en proyectos de Machine Learning es la cantidad y calidad de los datos. Es decir, si los datos de los pozos de Bolivia son limitados o sesgados, esto podría afectar la capacidad del modelo para generalizar y aplicarse a otros contextos o regiones.

Limitaciones Técnicas

Las limitaciones técnicas pueden incluir la capacidad de cómputo disponible para entrenar el modelo. Si el hardware no es suficientemente potente, podría limitar la complejidad del modelo o la velocidad a la que se pueden realizar los cálculos, afectando la eficiencia del modelo.



Capítulo 2

MARCO TEÓRICO

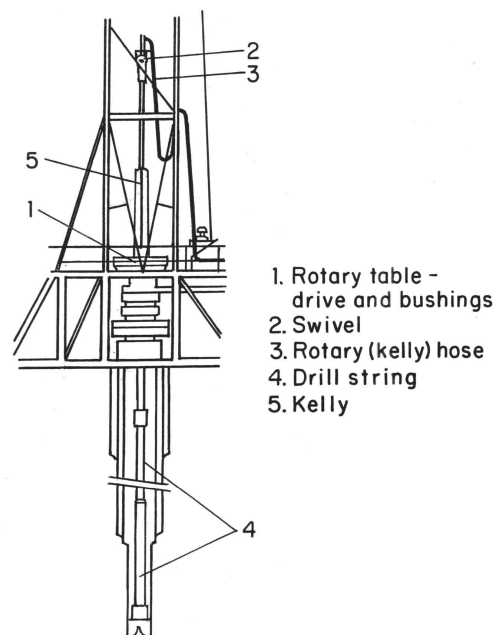
2.1. Conceptos Básicos de la Perforación de Pozos

La perforación de pozos es un procedimiento que básicamente implica atravesar el suelo y las formaciones adyacentes mediante un equipo de perforadora. El diámetro y la profundidad del pozo dependen de la necesidad, la disponibilidad hídrica, la geología presente entre otros. En cuanto a la ingeniería de proyecto está compuesta el equipo de ingeniería de perforación que proporciona soporte técnico para que la perforación sea óptima, incluyendo la selección y diseño del programa de lodo, los programas de revestimiento y cementación, el programa hidráulico, el programa de trepanos, el programa de sarta de perforación y el programa de control de pozos. Una vez que la perforación comienza, las operaciones diarias son manejadas por el equipo de operaciones de perforación, que incluye un supervisor y varios equipos de perforación (personal de torres de perforación y motores, perforadores, etc.) (Azar & Samuel, 2007).

El equipo de una plataforma de perforación rotativa, como se muestra en la figura 2.1, se compone de los siguientes sistemas:

- Sistema de generación de energía: proporciona la energía necesaria para operar los demás sistemas.
- Sistema de elevación: levanta la sarta de perforación y el trépano fuera del pozo.
- Sistema de circulación de fluido de perforación: circula el fluido de perforación a través del pozo para limpiar los recortes de perforación y controlar la presión de la formación.
- Sistema rotativo: rota la sarta de perforación y el trépano para perforar el pozo.
- Sistemas de control de reventón de pozos: protegen el pozo contra la presión de la formación.
- Sistema de adquisición y monitoreo de datos de perforación: recopila y registra datos de perforación para su análisis.

Figura 2.1: Sistema de Rotación.



Fuente: Adaptado de *Drilling Engineering* por Azar y Samuel (2007).

2.2. Tipos de Pozos

2.2.1. Pozos de Investigación Estratigráfica

Un pozo estratigráfico es una perforación geológica de investigación que tiene como objetivo la obtención de información específica del subsuelo, mediante la toma de registros geofísicos y el muestreo de la columna estratigráfica, es decir, de los distintos estratos de rocas. Este tipo de pozo se perfora con propósitos de reconocimiento y muestreo, sin el objetivo de producir hidrocarburos. Su profundidad suele estar entre 100 y 1000 metros, aunque puede ser mayor. Se utiliza para determinar la secuencia litológica y las propiedades petrofísicas y geoquímicas de la columna estratigráfica existente en el subsuelo. Los pozos estratigráficos son importantes para la exploración geológica y la identificación de condiciones propicias para el hallazgo de prospectos para hidrocarburos, yacimientos minerales, desarrollo de obras civiles, entre otros.

2.2.2. Pozos Exploratorios

Un pozo exploratorio se caracteriza por su función de investigación en nuevas áreas de acumulación de hidrocarburos, específicamente en zonas previamente no exploradas en busca de petróleo o gas. Estos pozos pueden emplazarse tanto en territorios nuevos como en secciones aún no exploradas dentro de un yacimiento existente. Es esencial notar que, inicialmente, cualquier pozo destinado a la extracción de hidrocarburos se considera exploratorio. Sin embargo, tras su construcción, su clasificación puede variar dependiendo del lugar donde se perforó y los resultados obtenidos, pudiendo convertirse en productores o mantenerse como no productores (EnergyAPI, 2001).

2.2.3. Pozos de Desarrollo

Los pozos de desarrollo se caracterizan por su finalidad de explotar y extraer las reservas presentes en un yacimiento. Estos pozos son perforados con el objetivo principal de incrementar la producción del campo y, por lo tanto, se ubican dentro del área previamente probada. A pesar de ello, debido a la incertidumbre en torno a la estructura o delimitación de los yacimientos, algunos pozos de desarrollo podrían no alcanzar sus objetivos previstos (EnergyAPI, 2001).

2.2.4. Pozos de Avanzada

Los pozos de avanzada surgen posteriormente a la perforación exitosa de un pozo exploratorio en una zona previamente no explorada que resulta ser productiva. La función principal de estos pozos es delimitar los límites del yacimiento. Sin embargo, también se emplean para ampliar el área conocida de un yacimiento, basándose en información obtenida durante la explotación, que sugiere una posible extensión más allá de los límites previamente estimados. Estos pozos, al estar ubicados fuera del área probada, conllevan un mayor nivel de riesgo en comparación con los pozos de desarrollo debido a su ubicación y la incertidumbre asociada (EnergyAPI, 2001).

2.3. Perforación de Pozos

El objetivo de la perforación es construir un pozo útil; un conducto desde el yacimiento hasta la superficie, que permita su explotación racional en forma segura y al menor costo posible (Mitchell et al., 2011).

2.3.1. Coordenadas del Conductor y Objetivos

El área escogida para perforar es producto de los estudios geológicos y/o geofísicos. La intención primordial de estos estudios es evaluar las excelentes, buenas, regulares o negativas perspectivas de las condiciones geológicas del subsuelo para emprender o no con el equipo de perforación la verificación de nuevos campos petrolíferos (Azar & Samuel, 2007).

2.3.2. Posición Estructural

El primer paso en la planeación de un pozo es la recolección de información de los pozos vecinos perforados en el área (no aplica a pozos exploratorios), una vez que se establecen los objetivos del pozo, se deberán considerar los pronósticos geológicos que consisten en:

- La columna estratigráfica esperada.
- La identificación de las anomalías geológicas que puedan encontrarse durante la perforación del pozo.

En la mayoría de los casos se obtiene de primera mano, un plano de ubicación y un plano que muestra las características de la estructura, el cual conforma el área de interés acompañado generalmente de una sección diagramática que muestra el perfil del objetivo con respecto a pozos vecinos (Chilingar et al., 2002).

2.3.3. Sarta de Perforación

La sarta de perforación es una columna de elementos metálicos armados de forma secuencial, de fabricación y especificaciones especiales, está conformado por el BHA y por la tubería de perforación. En el extremo inferior del BHA va enroscada el trepano, pieza también de fabricación y especificaciones especiales, que se

encarga de perforar los estratos geológicos para hacer el hoyo que llegará al yacimiento.

La sarta de perforación debe cumplir las siguientes funciones:

- Proporcionar Profundidad Vertical Verdadera (TVD).
- Conducir el fluido en su ciclo de circulación.
- Darle verticalidad o direccionalidad al pozo.
- Proteger la tubería del pandeo y de la torsión.
- Reducir el dog leg (pata de perro).
- Asegurar la bajada de la cañería.
- Reducir daño por vibración al equipo de perforación.
- Servir como herramienta complementaria de pesca.

El BHA es el arreglo de componentes mecánicos y eléctricos que van en la parte inferior de la sarta de perforación, generalmente consiste; del trepano, un motor de fondo (en ciertos casos), estabilizadores, el portamecha o Drill Collar, la columna de perforación pesada, los dispositivos que operan por golpes (tijeras o percutores), los Crossovers para las diversas formas de roscas y herramientas de medición durante la perforación como los Measuring While Drilling (MWD) y Logging While Drilling (LWD) (Jeffery, 2015). El arreglo de fondo de pozo debe proporcionar la fuerza para que el trepano fracture la roca (peso sobre el trepano), sobreviva en un ambiente mecánico hostil y proporcione al perforador el control direccional del pozo (Short, 1993).

2.3.4. Parámetros de Perforación

Los parámetros de perforación de pozos petroleros constituyen un conjunto de variables fundamentales que influyen en el proceso de perforación y en la obtención

exitosa de hidrocarburos. Estos parámetros son cruciales para la seguridad, la eficiencia y el rendimiento del pozo. Algunos de los parámetros claves incluyen:

- Profundidad del Pozo
- Presión de la bomba
- Tipo de Formación Geológica
- Diámetro del Pozo
- Temperatura y Presión de Formación
- Tasa de Penetración
- Tipos de Fluidos de Perforación
- Métodos de Control de Sólidos
- Peso sobre el trépano (WOB)
- Tipo de trépano de Perforación
- Desviación del Pozo

Profundidad del Pozo

La profundidad del pozo en la industria petrolera es un factor crítico que influye significativamente en las operaciones de perforación, exploración y producción de hidrocarburos. Este parámetro se refiere a la medida vertical u horizontal desde la superficie terrestre hasta el punto de interés en el subsuelo, determinado por los objetivos de la perforación y las características geológicas del área.

La determinación de la profundidad del pozo se basa en diversos factores, como el tipo de formaciones geológicas a perforar, los objetivos exploratorios o de producción, y las estrategias de extracción de hidrocarburos. Los pozos pueden ser someros,

con profundidades menores a cientos de metros, utilizados para exploración o extracción en formaciones poco profundas, o pueden ser pozos ultra profundos, con profundidades que pueden alcanzar varios miles de metros, necesarios para alcanzar reservorios más profundos o complejos.

La profundidad del pozo es un componente esencial en la planificación y ejecución de las operaciones de perforación. Factores como la presión del subsuelo, la temperatura, la estabilidad del pozo y la selección de equipos y tecnologías de perforación están directamente influenciados por la profundidad a la que se perfora el pozo.

La información precisa sobre la profundidad del pozo es crucial para diseñar estrategias de perforación efectivas, evaluar la viabilidad económica de la extracción de hidrocarburos y mitigar riesgos asociados con la perforación, como las patadas de pozo o los reventones. Además, la profundidad del pozo también puede influir en los costos operativos y en la planificación de infraestructuras relacionadas, como las plataformas petroleras (Adams & Charrier, 1985).

Presión de las Bombas

La presión de la bomba es un parámetro crítico en la industria petrolera que desempeña un papel fundamental en la seguridad, la eficiencia y la productividad de las operaciones de perforación y producción de hidrocarburos. Se refiere a la presión ejercida por los fluidos en el interior del pozo durante la perforación, y es esencial para mantener el control y la estabilidad del mismo.

La comprensión y el control adecuado de la presión de pozo son esenciales para prevenir situaciones de riesgo como patadas de pozo y reventones, que pueden representar peligros significativos para los trabajadores en la plataforma y ocasionar daños al equipo. Para mantener un equilibrio adecuado de presión, se utilizan fluidos de perforación específicos y sistemas de control, como los preventores de reventón

BOP, que sellan el pozo en caso de emergencia.

La presión de pozo varía a medida que se perfora a diferentes profundidades y a través de distintas formaciones geológicas. Los ingenieros de perforación monitorean constantemente la presión y ajustan los sistemas de control para garantizar que se mantenga dentro de límites seguros, evitando fluctuaciones bruscas que puedan desestabilizar el pozo.

Además de mantener la seguridad, la presión de pozo también afecta la eficiencia y la productividad de la extracción de hidrocarburos. Un control preciso de la presión permite maximizar la tasa de penetración y optimizar la producción de petróleo y gas de manera segura y rentable (Adams & Charrier, 1985).

Tipos de Formación Geológica

Una formación geológica es una unidad litoestratigráfica formal que define cuerpos de rocas caracterizados por propiedades litológicas comunes (composición y estructura) que las diferencian de las adyacentes.

Los tipos de formaciones geológicas son elementos fundamentales en la industria de la perforación de pozos petroleros, ya que influyen en la planificación, ejecución y evaluación de la viabilidad de la extracción de hidrocarburos. Estas formaciones se clasifican en base a sus características físicas, composición y capacidad para albergar reservorios de petróleo y gas.

Formaciones Sedimentarias Son las más comunes y se componen de sedimentos depositados por la acción del agua, viento o hielo. Ejemplos incluyen areniscas, lutitas y calizas. Estas formaciones suelen albergar los principales reservorios de hidrocarburos.

Formaciones Igneas Se originan por el enfriamiento y solidificación del magma o la lava. A menudo, estas formaciones no son consideradas reservorios importantes, aunque pueden afectar la perforación debido a su dureza.

Formaciones Metamórficas Resultan de la transformación de otras rocas debido a altas presiones y temperaturas. Ejemplos son el gneis y el esquisto metamórfico. Generalmente, estas formaciones no albergan grandes reservorios de hidrocarburos.

La comprensión de la formación geológica es crucial para determinar la viabilidad y el rendimiento de la extracción de petróleo y gas. La identificación precisa de la formación a perforar es esencial para seleccionar las herramientas, técnicas y fluidos de perforación adecuados, minimizando riesgos y maximizando la eficiencia durante la operación de perforación (Chilingar et al., 2002).

Diámetro del Pozo

El diámetro del pozo es un factor crítico en la industria petrolera, determinando el tamaño del agujero perforado en la formación geológica durante las operaciones de perforación. Esta dimensión, variando desde estrechos a anchos dependiendo de los objetivos de la perforación, influye en aspectos operativos y técnicos. Además de afectar el diseño de la cañería y la capacidad de flujo de los fluidos, impacta en aspectos económicos y ambientales, como costos de construcción y mantenimiento, así como la estabilidad estructural del pozo durante la perforación y la posterior producción. La elección del diámetro del pozo resulta crucial para garantizar la integridad y minimizar riesgos técnicos en la extracción de hidrocarburos.

El diámetro del pozo es un factor determinante en la planificación y ejecución de las operaciones de perforación en la industria petrolera. Su selección adecuada impacta en múltiples áreas técnicas, económicas y de seguridad. Variando desde dimensiones estrechas hasta más anchas, esta medida influye en el diseño de la

tubería, la capacidad de flujo de los fluidos, costos de construcción y la estabilidad estructural del pozo. Una elección acertada resulta fundamental para asegurar la integridad del pozo y minimizar riesgos técnicos y económicos asociados a la perforación y producción de hidrocarburos (Azar & Samuel, 2007).

Temperatura y Presión de Formación

La temperatura y presión de formación son elementos cruciales en la industria petrolera, refiriéndose a las condiciones térmicas y de presión existentes en las capas geológicas subterráneas donde se encuentran los yacimientos de hidrocarburos. Estos factores, determinados por la profundidad del yacimiento, influyen significativamente en la viabilidad de la extracción de petróleo y gas, así como en el diseño de las operaciones de perforación y producción. La comprensión precisa de la temperatura y presión de formación es esencial para determinar la estabilidad del pozo, la selección de equipos adecuados, la elección de fluidos de perforación y el diseño de estrategias para la extracción segura y eficiente de los recursos petrolíferos (Azar & Samuel, 2007).

Tasa de Penetración ROP

La Tasa de Penetración (ROP, por sus siglas en inglés) se refiere a la velocidad a la que avanza el trépano de perforación a través de las formaciones rocosas durante las operaciones de perforación. Es un parámetro crucial en la perforación, ya que influye directamente en la eficiencia y duración del proceso de perforación. Los ingenieros de perforación y los operadores supervisan de cerca la ROP, buscando optimizarla y considerando factores como el tipo de roca que se perfora, el diseño del trépano, las propiedades del fluido de perforación y el equipo utilizado. A menudo se busca alcanzar ROP más altas, ya que indican un progreso más rápido, pero lograr esto debe equilibrarse con factores como limitaciones del equipo y la estabilidad

del pozo.

Control de Sólidos

Los equipos de control de sólidos representan un conjunto de técnicas y dispositivos utilizados en la industria petrolera para procesar los recortes de roca generados durante la perforación de pozos. Estos sistemas, que incluyen zarandas, centrifugadoras y agitadores, separan los recortes sólidos del fluido de perforación, permitiendo su recirculación y manteniendo la estabilidad y calidad del fluido. La efectividad de estos equipos es crucial para prevenir obstrucciones en el pozo, optimizar la eficiencia de la perforación y minimizar el desgaste del equipo, contribuyendo así a operaciones de extracción de hidrocarburos seguras y eficientes (EnergyAPI, 2001).

Peso sobre el Trépano WOB

El peso sobre el trépano (WOB, por sus siglas en inglés) constituye un parámetro crítico en la industria petrolera, representando la fuerza aplicada sobre el trépano de perforación durante la operación. Este peso, controlado y ajustado por los operadores de la plataforma, se utiliza para facilitar la penetración en la formación geológica y optimizar la eficiencia de la perforación. El WOB se adapta según las características de la formación, la resistencia de la roca y la potencia requerida, buscando alcanzar un equilibrio que permita una perforación efectiva sin dañar el trépano ni comprometer la integridad del pozo. Un control preciso del peso sobre el trépano es esencial para maximizar la tasa de penetración y garantizar operaciones de perforación seguras y eficientes en la extracción de hidrocarburos (EnergyAPI, 2001).

Tipo de Trépano de Perforación

La selección del tipo de trépano de perforación en la industria petrolera es una decisión estratégica clave que impacta directamente en la eficiencia y éxito de las operaciones de perforación. Los trépanos de triconos, Diamond Compact Bits (PDC) y otras variantes, son diseñadas para adaptarse a diferentes tipos de formación geológica, ofreciendo capacidades de perforación específicas. La elección del tipo de trépano se basa en consideraciones como la dureza y composición de la roca, la profundidad del pozo, y las condiciones operativas. Cada tipo de trépano tiene sus ventajas y limitaciones, y seleccionar la más adecuada es esencial para optimizar la tasa de penetración, minimizar el desgaste del equipo y garantizar una extracción segura y eficiente de los recursos hidrocarburíferos (EnergyAPI, 2001).

Desviación del Pozo

La desviación del pozo es un término relevante en la industria petrolera que describe el grado en que un pozo perforado se aleja de su trayectoria vertical. Esta desviación puede ser intencional, utilizando técnicas de dirección, o incidental, debido a la naturaleza de las formaciones geológicas. La desviación controlada se emplea para alcanzar objetivos específicos, como llegar a reservorios distantes desde una sola ubicación o evitar áreas problemáticas. Sin embargo, un exceso de desviación no planificada puede aumentar la complejidad operativa y el tiempo de perforación. La gestión precisa de la desviación del pozo es esencial para lograr los objetivos de perforación de manera eficiente, garantizando la seguridad y maximizando la recuperación de hidrocarburos.

2.4. Fluidos de Perforación

Un Fluido de Perforación es una mezcla homogénea entre un fluido base (agua, aceite o gas) y diversos aditivos con propiedades fisicoquímicas específicas, que le permiten desempeñar diversas funciones durante las operaciones de perforación.

2.4.1. Funciones del Fluido de Perforación

Las funciones del fluido de perforación describen las tareas que el fluido de perforación es capaz de desempeñar, aunque algunas de éstas no sean esenciales en cada pozo. La remoción de los recortes del pozo y el control de las presiones de la formación son funciones sumamente importantes (EnergyAPI, 2001). Las funciones más importantes del fluido de perforación son las siguientes:

- Retirar los recortes del pozo.
- controlar las presiones de formación.
- Suspender y mantener en suspensión.
- Obturar las formaciones permeables.
- Mantener la estabilidad del pozo.
- Minimizar los daños a la formación.
- Enfriamiento, lubricación del trépano y el conjunto de perforación.
- Transmitir la energía hidráulica a las herramientas.
- Asegurar una evaluación adecuada de la formación.
- Evitar la corrosión.
- Facilitar la cementación del revestimiento.

- Minimizar el impacto ambiental.

2.5. Balance de un Pozo

El balance de pozos, denominado *wellbore balancing* en inglés, constituye un aspecto esencial en la industria del petróleo y gas, enfocado en el mantenimiento y regulación de las presiones durante las operaciones de perforación. Este enfoque es crucial para asegurar la estabilidad del pozo, prevenir riesgos de patadas de pozo o reventones, y salvaguardar la seguridad durante las distintas etapas de perforación y producción. Su propósito fundamental radica en mantener una presión hidrostática adecuada en el pozo que supere la presión ejercida por los fluidos y gases en la formación geológica circundante. Este equilibrio se logra mediante el uso de fluidos de perforación que contrarrestan la presión subterránea, evitando la intrusión no deseada de fluidos y gases al pozo. Los ingenieros de perforación y operadores juegan un papel crucial en el balance de pozos, monitoreando y ajustando constantemente sistemas de control, como los preventores de reventón BOP, para mantener la presión del pozo dentro de límites seguros. Durante la perforación, se controla cuidadosamente la presión hidrostática en el pozo, asegurando que sea superior a la presión de la formación geológica circundante, para evitar cualquier flujo no controlado de fluidos o gases. La importancia del balance de pozos radica en la prevención de accidentes significativos y costosos. Mantener un control riguroso de la presión del pozo garantiza la seguridad de los trabajadores en la plataforma y minimiza los riesgos de incidentes, tales como patadas de pozo o reventones, que podrían impactar negativamente tanto en la seguridad como en la eficiencia de las operaciones (Hawker et al., 2001).

2.5.1. Perforación Bajo Balance

En este proceso, se controla la presión del pozo de manera que se mantiene menor que la presión de formación (la presión natural de los fluidos dentro de la formación geológica). Esto se logra mediante el uso de fluidos de perforación de baja densidad, nitrógeno o aire, que no ejercen una presión hidrostática superior a la presión de la formación.

El objetivo principal de la perforación bajo balance es evitar dañar la formación productiva durante la perforación del pozo. Al mantener la presión del pozo por debajo de la presión de formación, se minimiza el riesgo de pérdida de circulación de fluidos al reservorio y se reduce la posibilidad de dañar las rocas porosa y permeable que contienen los hidrocarburos.

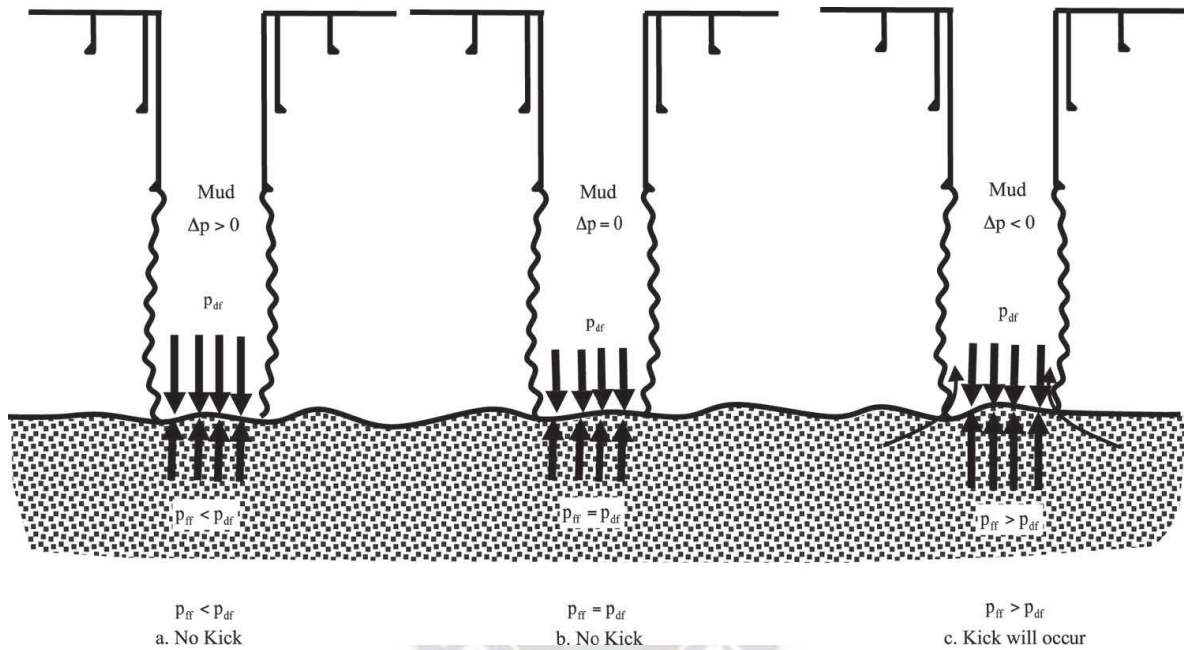
Esta técnica permite perforar pozos en reservorios con presiones bajas o frágiles sin perturbar la formación productora. Además, la perforación bajo balance puede resultar en una mejor productividad del pozo y reducir los costos operativos al minimizar los tiempos de detención para resolver problemas de formación (Hawker et al., 2001).

2.5.2. Perforación Sobre Balance

La perforación sobre balance es una estrategia utilizada en la industria petrolera donde la presión de los fluidos de perforación supera la presión de formación natural en la roca. Esta técnica busca mantener la estabilidad y prevenir la entrada no deseada de fluidos del yacimiento al pozo durante la perforación. Se emplean fluidos con una presión hidrostática mayor a la presión de la formación, evitando que los fluidos del reservorio fluyan hacia el pozo. Aunque esta técnica puede prevenir problemas de estabilidad del pozo, como el colapso de las paredes, y evitar la entrada de fluidos indeseados, como el agua o gas, también puede plantear desafíos,

como el riesgo de dañar la formación productiva o pérdidas de fluido. En contextos específicos, la elección de perforación sobre balance debe considerarse con precaución, evaluando sus ventajas y desventajas en función de las condiciones geológicas y operativas para garantizar una perforación segura y eficiente (Hawker, 2001).

Figura 2.2: Presion Diferencial.



Fuente: Adaptado de *Análisis de las presiones anormales de formación* por Hawker (2001).

2.6. Control de Pozos

2.6.1. Presión

La presión se define como la fuerza ejercida por unidad de área por parte de un fluido, típicamente medida en libras por pulgada cuadrada (psi). En la industria petrolera, se trabajan con distintos tipos de presión, entre ellos, la Presion de Formacion (PFm), Presion Hidrostatica (PH), Presion de Fricción (PDFr) y Presion de Fractura (PDF) (Hammack et al., 2010).

2.6.2. Presión de Fluido

Los fluidos son sustancias con capacidad de fluir. Inmediatamente se asocia la condición de “fluido” con líquidos como agua y petróleo, así como con gases. Para nuestros propósitos, nos referimos a los fluidos comúnmente utilizados en la industria petrolera, tales como petróleo, gas, agua, fluidos de terminación, agua salada, lodos, fluidos de empaque, entre otros, a menos que se especifique lo contrario. La presión ejercida por un fluido es el resultado de su densidad y la altura de la columna del mismo. En la industria petrolera, la densidad se mide en libras por galón (lpg; pounds per gallon; ppg) (gramo por litro, g/l). Un fluido mayor densidad producirá mayor presión debido a su peso. La fuerza se cuantifica en libras por pulgada cuadrada (lbs/plg²) (pounds per square inch, psi) (bar). Se calcula la presión que ejerce un fluido de densidad específica utilizando el gradiente de presión, expresado como la fuerza ejercida por pie de altura, medido en psi/pie (libras por pulgada cuadrada por pie) (bar/10 m). Para esta medida, se convierte la densidad del fluido de libras por galón (lpg; ppg) a psi/pie (o de g/l a bar/10m para bar/10m).

2.6.3. Presión Hidrostática

La presión hidrostática es la presión ejercida por un líquido en reposo debido a la fuerza gravitatoria que actúa sobre él. Se genera cuando un fluido está en un recipiente y su propio peso crea una presión en la parte inferior del contenedor y en cualquier objeto sumergido en él. Esta presión aumenta con la profundidad debido al incremento en la cantidad de líquido que hay por encima del punto considerado. La fórmula para calcular la presión hidrostática es:

$$P = \rho \cdot g \cdot h \quad (2.1)$$

donde:

P es la presión hidrostática en libras por pulgada cuadrada (psi).

ρ es la densidad del líquido en libras por galón (ppg).

g es la aceleración debido a la gravedad en pies por segundo cuadrado (ft/s²).

h es la altura o profundidad del líquido en pies (ft).

Este tipo de presión es fundamental en varios contextos, como en la determinación de la presión en profundidades marinas o en la comprensión de cómo funcionan los sistemas hidráulicos (White, 2016).

2.6.4. Presión de Formación

La presión de formación tiene su origen en los líquidos o gases presentes en los huecos de la estructura subterránea. Esta presión puede variar debido al peso ejercido por las capas de roca sobre la formación, influyendo tanto en los líquidos como en los componentes sólidos de la estructura rocosa. Los sólidos que conforman la roca y los espacios entre ellos son fundamentales en este proceso. Si los líquidos pueden moverse libremente, los sólidos pueden perder soporte y compactarse entre sí, un fenómeno conocido como compactación. Las formas de calcular esta presión están relacionadas con la presión de los espacios entre los sólidos de la roca y la densidad de los líquidos que se encuentran en esos espacios. Por otro lado, el gradiente de sobrecarga describe cómo cambia la presión a medida que profundizamos, debido al peso de la roca junto con los líquidos o gases que contiene.

$$P_{\text{form}} = \nabla G_{\text{form}} \cdot TVD \quad (2.2)$$

2.6.5. Presión de Fractura

Los pozos, fundamentales en la industria extractiva, son estructuras perforadas en la tierra con el propósito de explorar y extraer recursos como petróleo, gas o agua. Su clasificación se divide entre pozos exploratorios, destinados a investigar nuevas acumulaciones de hidrocarburos, y pozos productores, que buscan la extracción de estos recursos ya descubiertos. La presión en estos pozos desempeña un papel crucial: la presión de formación, generada por los fluidos en los espacios porosos de la formación rocosa, y la presión de fractura, indicadora de la fuerza necesaria para fracturar la estructura rocosa, son fundamentales para determinar la viabilidad y productividad del pozo. Comprender la presión en los pozos es esencial, ya que influye directamente en la capacidad de flujo del pozo y en su manejo. Estas presiones varían considerablemente con la profundidad y la geología del área. La presión de formación, por ejemplo, puede ser afectada por el peso de la roca sobre la formación, mientras que la presión de fractura indica cuánta fuerza se requiere para deformar permanentemente la estructura rocosa. Estos conceptos son cruciales para evaluar la viabilidad y el rendimiento de los pozos en la extracción de recursos naturales. El conocimiento de la presión en los pozos es esencial para la seguridad, la eficiencia y el éxito en la extracción de recursos. La comprensión de estas presiones no solo influye en la productividad del pozo, sino también en las estrategias de extracción y la viabilidad económica de la operación. La variabilidad de estas presiones con la geología y la profundidad subraya la necesidad de análisis detallados y precisos en la industria extractiva para garantizar una extracción eficiente y segura de recursos (EnergyAPI, 2001).

2.6.6. Presión de Fondo de Pozo

La presión de fondo de pozo (BHP) (*del inglés Bottom Hole Pressure*), es la presión que se registra en el fondo del pozo, ya sea por la presencia de fluidos de formación (petróleo, gas, agua) o por el lodo de perforación utilizado durante la perforación. Medida generalmente en libras por pulgada cuadrada (psi). Esta presión puede ser calculada en un pozo estático lleno de fluido con la ecuación:

$$BHP = MW \cdot TVD \cdot 0,052 \quad (2.3)$$

donde:

BHP es la presión de fondo de pozo en libras por pulgada cuadrada (lb/in²).

MW es el peso del lodo en libras por galón (lpg).

TVD es la profundidad vertical verdadera en pies (ft).

0,052 es un factor de conversión si se utilizan estas unidades de medida.

En los pozos en circulación, el incremento de la BHP es equivalente a magnitud de la fricción del fluido en el espacio anular. El gradiente de BHP debe exceder el gradiente de presión de formación para evitar un influjo del fluido de formación en el pozo. Por otra parte, si la BHP (incluida la presión adicional de fricción del fluido de un fluido en proceso de flujo) es demasiado alta, una formación débil puede fracturarse y causar una pérdida de fluidos del pozo. La pérdida de fluido en una formación puede ser seguida por el influjo de fluido desde otra formación (EnergyAPI, 2001).

2.6.7. Densidad Equivalente de Circulación

Durante el proceso de circulación, la presión en el fondo del pozo y en todo el espacio anular aumenta debido al movimiento del fluido. Este incremento de presión es resultado de la fricción del fluido y de las pérdidas de presión que ocurren en el espacio anular mientras el fluido se desplaza. Entender esta presión es crucial durante la perforación, ya que la presión que equilibra el pozo es mayor en este momento en comparación con la presión ejercida por la columna estática de lodo. Una mayor presión de circulación conducirá a:

- Mayor sobrebalance en comparación con la presión de formación
- Mayor riesgo de inundación de la formación
- Invasión más grave de la formación
- Mayor riesgo de atascamiento diferencial
- Mayor carga ejercida sobre el equipo en superficie

La mayor presión se denomina “presión dinámica” ó “presión circulante de fondo” (*BHCP*, por sus siglas en inglés: *Bottom Hole Circulating Pressure*).

$$BHCP = P_{\text{lodo}} + \Delta Pa \quad (2.4)$$

donde ΔPa es la suma de las pérdidas de presión anular.

Cuando esta presión se convierte al peso equivalente de lodo, se utiliza el término “densidad equivalente de circulación” (*ECD*, por sus siglas en inglés: *Equivalent Circulating Density*).

$$ECD = MW + \frac{\Delta Pa}{(g \cdot TVD)} \quad (2.5)$$

El peso de los restos o desechos de la perforación también es relevante durante el proceso de perforación. Estos residuos, presentes en el espacio anular en cualquier

etapa, se suman al peso del lodo, incrementando la presión en el fondo del pozo. Similar a lo que ocurre con el aumento de la presión de fondo durante la circulación ECD, se evidencian cambios de presión debido al movimiento inducido del lodo y las consiguientes presiones de fricción al introducir o sacar la tubería del pozo (Hawker, 2001).

2.7. Parámetros de Perforación e Influencia en la ROP

La mejora de la eficiencia en la perforación es crucial, especialmente cuando los precios del petróleo están en declive. La Rate of Penetration (ROP) desempeña un papel fundamental en este proceso. Durante las etapas de planificación y exploración de pozos, la elección de trépanos y parámetros de perforación afecta significativamente el costo total y la duración de la operación. Sin embargo, optimizar la ROP es un desafío debido a la compleja relación entre las variables de perforación y la tasa de penetración. Para abordar este problema, se requiere el desarrollo de sistemas informáticos avanzados, modelos predictivos y algoritmos eficientes.

2.8. Modelos de ROP

2.8.1. Modelos Mecánicos

Modelo de Teale

El modelo de Teale es un modelo mecánico para predecir la Rate of Penetration (ROP) en la perforación rotatoria. Se basa en la teoría de la mecánica de rocas y fue desarrollado por R.L. Teale en 1965. El modelo de Teale asume que la ROP es proporcional a la potencia disponible en el trépano y a la inversa de la resistencia a la penetración de la formación. La potencia disponible en el trépano se calcula

como el producto del peso sobre el trépano (WOB) y la velocidad de rotación (RPM).

La resistencia a la penetración de la formación se calcula como la suma de la resistencia a la compresión de la formación (UCS) y la resistencia a la cizalladura de la formación.

La ecuación del modelo de Teale para la ROP es la siguiente:

$$ROP = \frac{K \cdot WOB \cdot RPM}{UCS + \tau} \quad (2.6)$$

donde:

ROP Tasa de penetración (ft/hr)

K Constante de proporcionalidad (depende del trépano y la formación)

WOB Peso sobre el trépano (lb)

RPM Velocidad de rotación (rev/min)

UCS Resistencia a la compresión de la formación (psi)

τ Resistencia a la cizalladura de la formación (psi)

El modelo de Teale es un modelo simplificado y no tiene en cuenta todos los factores que afectan la ROP, algunas de las limitaciones del modelo incluyen:

- No tiene en cuenta el efecto del tipo de trépano.
- No tiene en cuenta el efecto del caudal de lodo.
- No tiene en cuenta el efecto de las propiedades de la formación como la porosidad y la permeabilidad.

El modelo de Teale se puede utilizar para:

- Predecir la ROP para diferentes parámetros de perforación.
- Optimizar los parámetros de perforación para maximizar la ROP

- Seleccionar el trépano adecuada para una formación específica.

Modelo de Bingham

El modelo de Eugene Clarence Bingham es un modelo reológico que se utiliza para describir el comportamiento de fluidos no newtonianos como el lodo de perforación, el cual se basa en la idea de que el lodo de perforación tiene un límite de fluencia, que es la tensión mínima que se debe aplicar para que el fluido comience a fluir. Por otra parte Michael G. Bingham basándose en el modelo de E. C. Bingham, modificó el modelo de Maurer al ignorar la profundidad de perforación, considerando solo en función de la WOB, RPM y el diámetro, e introdujo dos nuevas constantes a y k ambos adimensionales (Bingham, 1965).

$$ROP = k \cdot \left(\frac{WOB}{d} \right)^a \cdot RPM \quad (2.7)$$

ROP es la tasa de penetración (ft/h)

K es la constante de perforabilidad (Klbf)

WOB peso sobre el trépano (Klbf)

a parámetro de ajuste que representan una medida de la capacidad de perforación de cada litología (exponente del peso del trépano)

d diametro del trépano (inch)

RPM es la revolución por minuto (rpm)

El modelo de M. G. Bingham es un modelo simplificado y no tiene en cuenta todos los factores que afectan la ROP. Algunas de las limitaciones del modelo incluyen:

- No tiene en cuenta el efecto de la temperatura.
- No tiene en cuenta el efecto de la presión.

- No tiene en cuenta el efecto de la profundidad.
- No tiene en cuenta el efecto de la composición del lodo de perforación.
- Esta limitado a valores bajos de RPM y WOB.

2.8.2. Modelos Empíricos

Modelo de Maurer

Maurer propuso un modelo teórico para trépanos de cono de rodillo basado en varios parámetros: peso sobre el trépano (WOB), tamaño del trépano, capacidad de perforación, resistencia de la roca (UCS) y velocidad de rotación. El modelo desarrollado por Maurer se muestra en la ecuación.

$$ROP = K \cdot \frac{RPM \cdot WOB^2}{d^2 \cdot UCS^2} \quad (2.8)$$

ROP es la tasa de penetración (rpm)

K es la constante de perforabilidad (Klbf)

WOB peso sobre el trépano (Klbf)

UCS resistencia a la compresión del trépano (KPa)

d diametro del trépano (in)

Modelo de Bourgoyne y Young

Bourgoyne y Young (1974) desarrollaron uno de los modelos ROP más importantes, que se emplea ampliamente en la industria. Según Soares y Gray (2019), el modelo de Bourgoyne y Young (BYM) incluye ocho parámetros y puede escribirse como:

$$ROP = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5 \cdot f_6 \cdot f_7 \cdot f_8 \quad (2.9)$$

donde f son las funciones:

f_1 capacidad de perforación de formación

f_2 disminución de la tasa de penetración con la profundidad

f_3 aumento de la penetración debido a la falta de compactación

f_4 cambio en la penetración debido al peso del lodo

f_5 cambio en la tasa de penetración debido al peso del lodo

f_6 cambio en la tasa de penetración debido al N (rpm)

f_7 cambio en la tasa de penetración debido al desgaste de los dientes

f_8 cambio en la tasa de penetración debido a la fuerza de impacto del chorro

Estas relaciones funcionales se pueden expresar de la siguiente manera:

$$f_1 = e^{(2,303 \cdot a_1)} \quad (2.10)$$

$$f_2 = e^{2,303 \cdot a_2 \cdot (10000 - TVD)} \quad (2.11)$$

$$f_3 = e^{2,303 \cdot a_3 \cdot D^{0,96} \cdot (\rho_f - 9)} \quad (2.12)$$

$$f_4 = e^{2,303 \cdot a \cdot TVD \cdot (\rho_f - \rho_m)} \quad (2.13)$$

$$f_5 = \left[\frac{\left(\frac{WOB}{d_b} \right) - \left(\frac{WOB}{d_b} \right)_t}{4 - \left(\frac{WOB}{d_b} \right)_t} \right]^{a_5} \quad (2.14)$$

$$f_6 = \left(\frac{RPM}{60} \right)^{a_6} \quad (2.15)$$

$$f_7 = e^{-a_7 \cdot h} \quad (2.16)$$

$$f_8 = \left(\frac{F_j}{1000} \right)^{a_8} \quad (2.17)$$

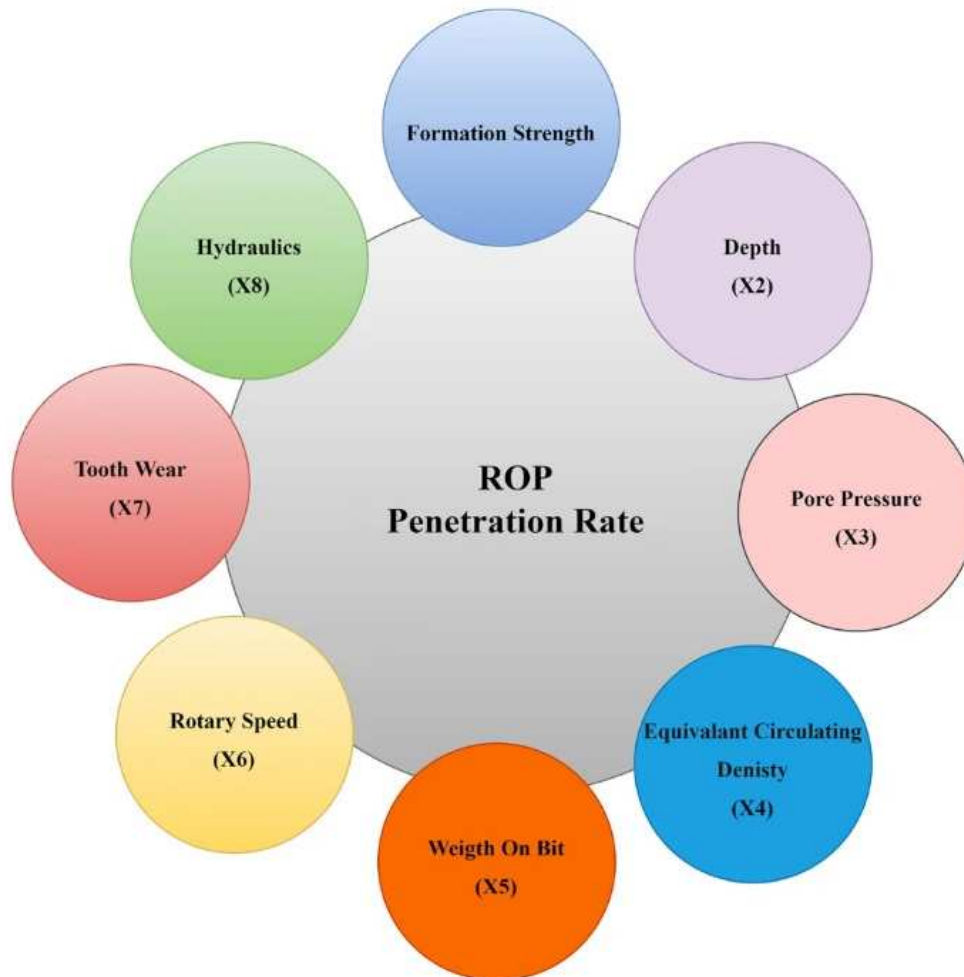
$$F_j = \rho q v_n \left(\frac{lbf}{32,174 \frac{ftlbm}{s^2}} \right) \left(\frac{mín}{60s} \right) = 0,000518 \rho q v_n \quad (2.18)$$

$$v_n = \frac{q}{TFA} \left(\frac{mín}{60s} \right) \left(\frac{ft^3}{7,4805gal} \right) \left(\frac{12in}{ft} \right)^2 = 0,3208 \frac{q}{TFA} \quad (2.19)$$

las constantes a_1 hasta a_8 deben elegirse en función de las condiciones locales de perforación

$$\frac{dD}{dt} = \exp\left(a_1 + \sum_{j=1}^8 a_j \cdot x_j\right) \quad (2.20)$$

Figura 2.3: Relación funcional de la tasa de penetración de acuerdo al modelo de Bourgoyne y Young.



Fuente: Adaptado de *Prediction and analysis of penetration rate in drilling operation using deterministic and metaheuristic optimization methods* por Sobhi et al. (2022).

2.8.3. Modelos de Regresión Lineal

Los modelos lineales son una herramienta poderosa para optimizar la ROP, estos modelos se basan en la regresión lineal, una técnica estadística que utiliza una ecuación lineal para predecir una variable dependiente (en este caso, la ROP) a partir de una o más variables independientes (como el peso sobre el trépano, la velocidad de rotación y el caudal de lodo).

Ejemplo de un modelo lineal simple:

$$ROP = a + b \cdot WOB \quad (2.21)$$

donde:

ROP Tasa de penetración (ft/hr)

a: Intercepto

b: Pendiente

WOB Peso sobre el trépano (lb)

Ejemplo de un modelo lineal múltiple:

$$ROP = a + b \cdot WOB + c \cdot RPM + d \cdot Q \quad (2.22)$$

donde:

ROP Tasa de penetración (ft/hr)

a Intercepto

b Coeficiente del peso sobre el trépano

c Coeficiente de la velocidad de rotación

d Coeficiente del caudal de lodo

WOB Peso sobre el trépano (lb)

RPM Velocidad de rotación (rev/min)

Q Caudal de lodo (gpm)

Los modelos lineales son una herramienta útil para optimizar la ROP, pero no son la única herramienta que se puede utilizar. Otros modelos más complejos, como las redes neuronales artificiales, pueden proporcionar una mayor precisión.

2.8.4. Modelos de Redes Neuronales

Una Red Neuronal Artificial (ANN) es una herramienta poderosa para optimizar la tasa de penetración (ROP) en la perforación de pozos. Las ANNs son modelos de aprendizaje automático que se inspiran en el funcionamiento del cerebro humano. Las ANNs pueden aprender de los datos y hacer predicciones precisas, incluso en situaciones complejas.

Ventajas de las ANNs para la optimización de la ROP:

- Precisión: Las ANNs pueden ser más precisas que los modelos tradicionales como los modelos lineales.
- Flexibilidad: Las ANNs pueden modelar relaciones no lineales entre las variables.
- Capacidad de aprendizaje: Las ANNs pueden aprender de los datos y mejorar su rendimiento con el tiempo.
- Robustez: Las ANNs son menos sensibles a los valores atípicos en los datos.

Desventajas de las ANNs para la optimización de la ROP:

- Complejidad: Las ANNs pueden ser más complejas y difíciles de entender que los modelos tradicionales.
- Requerimiento de datos: Las ANNs requieren una gran cantidad de datos para entrenar.
- Tiempo de entrenamiento: El entrenamiento de una ANN puede llevar tiempo.

Las ANNs son una herramienta poderosa para optimizar la ROP, las ANNs pueden ser más precisas que los modelos tradicionales y pueden modelar relaciones no lineales entre las variables. Sin embargo, las ANNs pueden ser más complejas y requerir una gran cantidad de datos para entrenar.

Tanto los modelos empíricos como mecánicos para la optimización de la ROP presentan desventajas como suposiciones simplificadas que no reflejan la complejidad real de la perforación, dependencia de datos históricos que pueden no ser representativos, limitación en las variables consideradas, dificultades para generalizar a diferentes condiciones, tendencia a soluciones de optimización local, consumo significativo de tiempo y recursos para su desarrollo e implementación, y falta de adaptabilidad en tiempo real. Estas limitaciones han llevado a la exploración de enfoques más avanzados como el ML y la AI, que ofrecen mejor manejo de grandes conjuntos de datos y adaptabilidad a condiciones cambiantes.

Por otro lado, los modelos tradicionales de ROP, como los de Bingham, Bourgoyne and Young, entre otros, a menudo se basan en variables petrofísicas que no siempre están disponibles en tiempo real o requieren cálculos complejos, lo que limita su utilidad en aplicaciones de predicción instantánea como se muestra en la tabla 2.1. Los modelos de ML, gracias a su capacidad para procesar y aprender de grandes volúmenes de datos operativos, pueden ofrecer predicciones más rápidas y precisas, lo que es esencial para la toma de decisiones en el campo.

Tabla 2.1: Variables de perforación para cuatro modelos analíticos de ROP

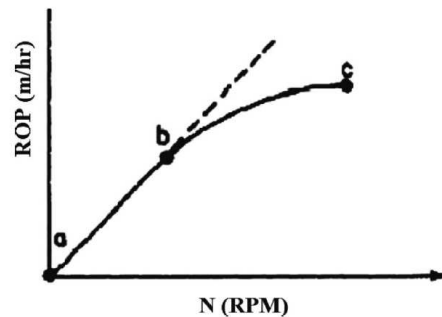
| Modelos de ROP | Bingham | Bourgoyne and Young | Hareland and Rampersad | Motahhari |
|--|----------------|----------------------------|-------------------------------|------------------|
| Variables de Operación | | | | |
| Peso sobre el trépano (WOB) | X | X | X | X |
| Revoluciones por minuto (RPM) | X | X | X | X |
| Flow Rate Q | | X | | |
| Profundidad Vertical Verdadera (TVD) | | X | | |
| Densidad Equivalente de Circulación (ECD) | | X | | |
| Bit Wear W_f | | X | | |
| Propiedades de la Roca | | | | |
| Gradiente de Presión de Poro | | X | | |
| Resistencia a la compresión no confinada | | | X | |
| Resistencia a la compresión confinada | | | | X |
| Propiedades del Trépano | | | | |
| Diámetro del Trépano | X | X | X | X |
| Diámetro de la Boquilla del Trépano | | X | | |
| Propiedades de Diseño de Cortadores | | | | |
| Número de cortadores | | | X | X |
| Diámetro del cortador | | | X | |
| Ángulo de rastrillo lateral del cortador | | | X | |
| Ángulo de rastrillo trasero del cortador | | | X | |
| Propiedades del Lodo | | | | |
| Mud Weigth (MW) | | X | | |
| Viscosidad Aparente a 10000 sec-1 | | X | | |

Fuente: Adaptado de *Computational prediction of the drilling rate of penetration (ROP): A comparison of various machine learning approaches and traditional models* por Brenjkar y Biniáz Delijani (2022).

2.9. Exponente D de Perforación

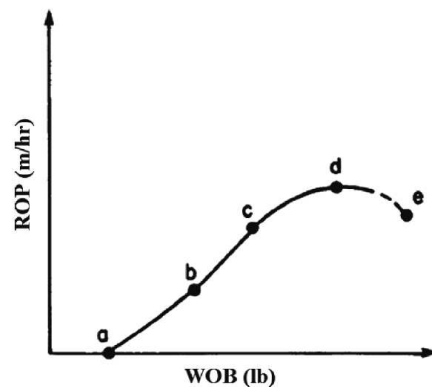
El exponente D y el exponente Dc corregido son parámetros utilizados en la predicción de geopresiones durante la perforación de pozos. Este exponente refleja la facilidad de perforación de una formación específica. A medida que la porosidad disminuye con la profundidad, la perforación se vuelve más difícil, lo que se traduce en un aumento del exponente D.

Figura 2.4: Rate of Penetration (ROP) Vs Revoluciones por minuto (RPM).



Fuente: Adaptado de *Origin and Prediction of Abnormal Formation Pressures* por Chilingar et al. (2002).

Figura 2.5: Rate of Penetration (ROP) Vs Peso sobre el trépano (WOB).



Fuente: Adaptado de *Origin and Prediction of Abnormal Formation Pressures* por Chilingar et al. (2002).

Exponente D Jordan y Shirley desarrollaron esta teoría en 1966, para derivar el modelo de M. G. Bingham obteniendo:

$$d = \frac{\log\left(\frac{ROP}{60 \cdot RPM}\right)}{\log\left(\frac{12 \cdot WOB}{10^6 \cdot D}\right)} \quad (2.23)$$

- Este método se aplica durante la perforación del pozo.
- Toma en cuenta varios factores, como el peso sobre el trépano, la velocidad de la mesa rotaria, la velocidad de penetración y el diámetro del trépano.
- Proporciona información sobre la perforabilidad de la formación.
- Un valor más alto de D indica una formación más fácil de perforar, mientras que un valor más bajo sugiere una formación más dura.

Exponente Dc El exponente Dc corregido, similar al exponente D, pero se corrige debido a que un cambio en el peso del lodo llevará a un cambio en la presión diferencial y como consecuencia, tendrá un efecto en el exponente d (2001). Es por ello que Rehm y McClendon, en 1971, desarrollaron la ecuación:

$$D_c = d \cdot \frac{\nabla G_{norm,form}}{MW} \quad (2.24)$$

D_c es el exponente D corregido

d es el exponente d

$\nabla G_{norm,form}$ es el gradiente de presión normal de formación

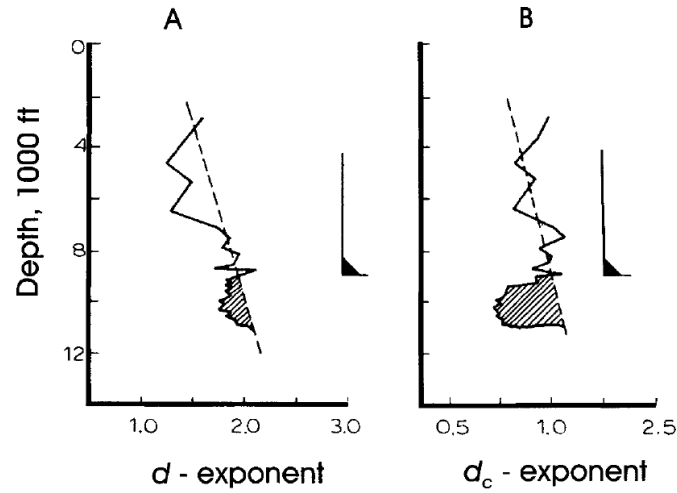
MW es el peso de lodo

El exponente Dc:

- Toma en cuenta la densidad del lodo, el pH, la viscosidad y la salinidad.
- Ayuda a ajustar la predicción de la presión de poro y la presión de fractura teniendo en cuenta las condiciones específicas del lodo utilizado durante la

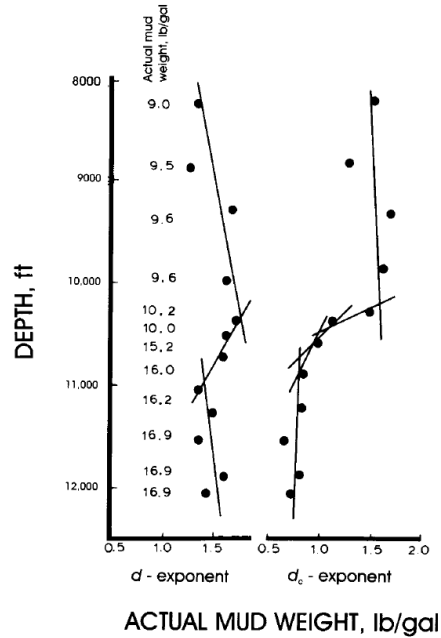
perforación.

Figura 2.6: Exponente D y Dc.



Fuente: Adaptado de *Origin and Prediction of Abnormal Formation Pressures* por Chilingar et al. (2002).

Figura 2.7: Comparación entre el exponente D y Dc.



Fuente: Adaptado de *Origin and Prediction of Abnormal Formation Pressures* por Chilingar et al. (2002).

Capítulo 3

ALGORITMOS DE MACHINE LEARNING PARA LA OPTIMIZACIÓN DE LA ROP

3.1. Machine Learning

El ML es una parte de la AI y la informática que se caracteriza por el uso de datos y/o algoritmos tal que se pueda interpretar la manera en la que aprenden los seres humanos (Education, 2020).

La AI y ML han crecido en popularidad a través de diversas industrias. Las corporaciones, las universidades, el gobierno y los grupos de investigación han notado el verdadero potencial de varias aplicaciones de AI y ML para automatizar varios procesos al tiempo que aumentan las capacidades de predicción. El potencial de la IA y el ML es un cambio de juego notable en varias industrias. Los avances tecnológicos de AI de los automóviles autónomos, la detección de fraudes, el reconocimiento de voz, el filtrado de spam, las recomendaciones de productos y contenido de Amazon

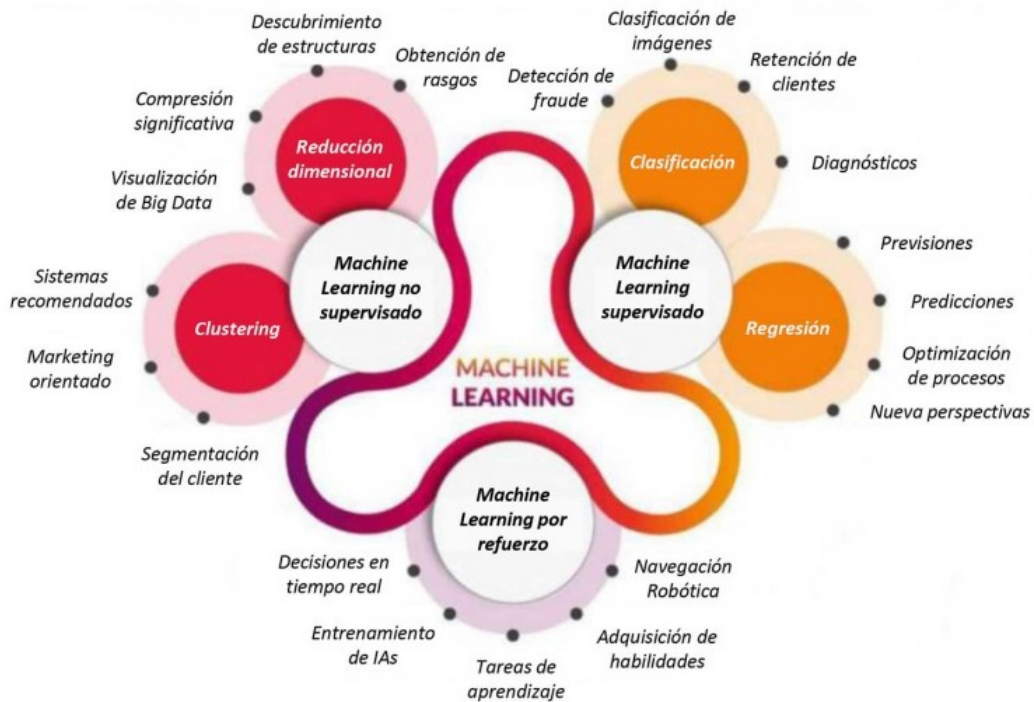
y Facebook, etc., han generado cantidades masivas de valor neto de activos para varias corporaciones. La industria de la energía se encuentra en la fase inicial de la aplicación de la AI a diferentes aplicaciones. El aumento de la popularidad en la industria de la energía se debe a las nuevas tecnologías como los sensores y los servicios informáticos de alto rendimiento (por ejemplo, Apache Hadoop, NoSQL, etc.) que permiten la adquisición y el almacenamiento de Big Data en diferentes campos de estudio. Big data se refiere a una cantidad de datos que es demasiado grande para ser manejado (es decir, recopilado, almacenado y analizado) utilizando herramientas y técnicas comunes, por ejemplo, terabytes de datos. El número de publicaciones en este ámbito ha aumentado exponencialmente a lo largo de los últimos años. A medida que más empresas se den cuenta del valor agregado a través de la incorporación de IA en las operaciones diarias, se fomentarán ideas más creativas (Belyadi & Haghighat, 2021).

La figura 3.1 proporciona una visión general de las tres categorías principales del Machine Learning: aprendizaje no supervisado, aprendizaje supervisado y aprendizaje por refuerzo.

El **Machine Learning** se está aplicando en la industria petrolera de varias maneras innovadoras, especialmente en la perforación de pozos. A continuación, se presentan algunos ejemplos de cómo se utiliza.

- **Automatización y Mantenimiento Predictivo:** Se utilizan sensores para recopilar datos y desarrollar modelos que anticipen fallas en las máquinas, permitiendo un mantenimiento preventivo.
- **Optimización de la Perforación:** El machine learning ayuda a optimizar el diseño, la perforación y la terminación de pozos, mejorando los volúmenes y costos de producción.
- **Análisis de Datos Sísmicos y Geológicos:** Se analizan grandes volúmenes

Figura 3.1: Aplicaciones del ML



Fuente: Adaptado de *El aprendizaje profundo y el machine learning generan escenarios de transformación educativos facilitados por procesos algorítmicos y neuronales artificiales* por Farnós (2023).

de datos sísmicos, registros de pozos y otros datos geológicos para identificar posibles yacimientos de petróleo y gas.

- **Predicción de Problemas Operativos:** Se emplean algoritmos para predecir problemas durante la perforación, lo que permite responder rápidamente a cualquier inconveniente que pueda surgir.
- **Modelado Predictivo:** Se implementan modelos predictivos de machine learning para estimar parámetros óptimos en la perforación, como la tasa de penetración y la eficiencia mecánica específica.

3.1.1. Machine Learning Supervisado

El aprendizaje supervisado es un enfoque en el campo del Machine Learning donde se entrena un modelo utilizando un conjunto de datos que incluye ejemplos de entrada y la salida esperada. En otras palabras, el modelo *aprende* a realizar predicciones o tomar decisiones basadas en ejemplos etiquetados. Los datos de entrenamiento se denominan *supervisados* porque el modelo recibe retroalimentación explícita durante el proceso de entrenamiento.

En el aprendizaje supervisado, el objetivo es que el modelo aprenda a mapear las entradas a las salidas deseadas, de modo que pueda generalizar y hacer predicciones precisas sobre datos no vistos previamente. Este enfoque se utiliza comúnmente en problemas de clasificación (donde se predice una etiqueta o categoría) y regresión (donde se predice un valor numérico) (Gil, 2020).

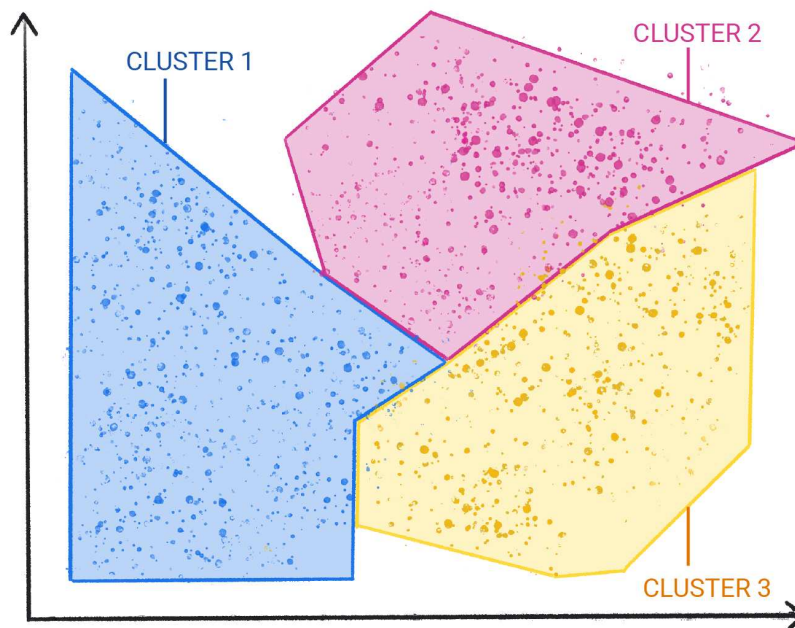
3.1.2. Machine Learning No Supervisado

El aprendizaje no supervisado es un tipo de aprendizaje automático o machine learning en el que los modelos aprenden a partir de datos sin etiquetas. En este tipo de aprendizaje, los modelos buscan encontrar patrones y relaciones en los datos sin tener acceso a información etiquetada que los permita clasificar o predecir. Algunas técnicas populares de aprendizaje no supervisado incluyen:

- Regla de asociación: Es un método de aprendizaje sin supervisión que se utiliza para encontrar patrones en los datos.
- Análisis de componentes principales PCA: Esta técnica se utiliza para reducir la dimensionalidad de los datos y encontrar patrones en ellos.
- Aprendizaje factorial: Es un enfoque que busca crear un código factorial de los datos, es decir, un código con componentes estadísticamente independientes.

El aprendizaje no supervisado es útil en diversas aplicaciones, como la identificación de patrones en grandes volúmenes de datos, la descubrimiento de grupos y la compresión de datos. A diferencia del aprendizaje supervisado, que requiere etiquetas de entrenamiento, el aprendizaje no supervisado se basa en la exploración de los datos sin etiquetas y permite identificar patrones y relaciones inesperadas. Sin embargo, el aprendizaje no supervisado puede ser menos preciso y confiable que el aprendizaje supervisado, ya que no se beneficia de la información proporcionada por las etiquetas de entrenamiento (IBM, 2024).

Figura 3.2: Grupos de clústeres con demarcaciones naturales.



Fuente: Adaptado de *Inteligencia Artificial: Fundamentos, Prácticas y Aplicaciones* por Serrano (2016).

3.1.3. Machine Learning por Refuerzo

El aprendizaje por refuerzo (RL, por sus siglas en inglés) es una rama del machine learning que se centra en la interacción entre un agente y un entorno. El agente busca tomar decisiones que maximicen las recompensas a lo largo del tiempo,

basándose en la evaluación de las acciones y estados en función de las recompensas que reciben. El aprendizaje por refuerzo se inspira en la psicología conductista y se ha aplicado con éxito en diversos campos, como el control de robots, telecomunicaciones y juegos.

El proceso de aprendizaje por refuerzo se puede dividir en los siguientes componentes principales:

Ambiente: Es el entorno en el que el agente se mueve e interactúa. Contiene todos los elementos que constituyen el estado y establece las reglas y limitaciones del sitio.

Estado: Es la situación actual del ambiente, que tiene en cuenta todos sus elementos y variables. En pocas palabras, es un indicador actual del ambiente.

Agente: Es el software que interactúa con el entorno y toma decisiones en función de las observaciones y las recompensas recibidas.

El aprendizaje por refuerzo se puede adaptar a diferentes problemas y aplicaciones, como la optimización de procesos, la planificación de rutas y la personalización de experiencias de usuario. Algunas técnicas populares de aprendizaje por refuerzo incluyen la fuerza bruta, Q-Learning y la ecuación de Bellman.

El aprendizaje por refuerzo es especialmente útil en situaciones donde se requiere adaptabilidad y capacidad de aprendizaje a partir de la experiencia, como en robótica, control de sistemas y optimización de recursos (AWS, 2022).

3.1.4. IA Generativa

La IA generativa es una clase de modelos que crea contenido a partir de entradas del usuario. Por ejemplo, la IA generativa puede crear imágenes novedosas, composiciones musicales y chistes; puede resumir artículos, explicar cómo realizar una tarea o editar una foto.

La IA generativa puede tomar una variedad de entradas y crear una variedad de salidas, como texto, imágenes, audio y video. También puede tomar y crear combinaciones de estos. Por ejemplo, un modelo puede tomar una imagen como entrada y crear una imagen y texto como salida, o tomar una imagen y texto como entrada y crear un video como salida.

Podemos analizar los modelos generativos por sus entradas y salidas, que por lo general se escriben como “tipo de entrada” a “tipo de salida”. Por ejemplo, la siguiente es una lista parcial de algunas entradas y salidas para modelos generativos:

- Texto a texto.
- Texto a imagen.
- Texto a video.
- Texto a código.
- Texto a voz.
- Imagen y texto a imagen.

La IA generativa es una tecnología que evoluciona rápidamente y cuenta con nuevos casos de uso que se descubren de forma constante. Por ejemplo, los modelos generativos ayudan a las empresas a definir mejor sus imágenes de productos de comercio electrónico, ya que quitan automáticamente fondos distractores o mejoran la calidad de las imágenes de baja resolución.

3.2. Herramientas y Librerías de Machine Learning

3.2.1. Python

Python es un lenguaje de programación informático que se utiliza para crear sitios web, software, automatizar tareas, análisis de datos y más. Es un lenguaje de propósito general, lo que significa que se puede utilizar para crear una variedad de programas diferentes y no está especializado en ningún problema específico. Python se ha convertido en uno de los lenguajes de programación más populares del mundo en los últimos años, utilizado en todo, desde el aprendizaje automático hasta la construcción de sitios web, y se utiliza por desarrolladores como no desarrolladores.

El término *Machine Learning con Python* se refiere al uso del lenguaje de programación Python como una herramienta principal para desarrollar y aplicar técnicas de Machine Learning. Python se ha convertido en un lenguaje muy popular en el campo del Machine Learning debido a su amplia gama de bibliotecas especializadas, su sintaxis clara y legible, y su comunidad activa de desarrolladores.

Al utilizar Python para Machine Learning, los profesionales y entusiastas de la ciencia de datos pueden aprovechar bibliotecas como NumPy, SciPy, Pandas, Matplotlib y scikit-learn, entre otras, para realizar tareas fundamentales como manipulación de datos, visualización, implementación de algoritmos de Machine Learning y evaluación de modelos (Oracle, 2023).

3.2.2. Librerías para Ciencia de Datos

SciPy

Es una biblioteca de código abierto para Python que proporciona herramientas y algoritmos matemáticos para aplicaciones en ciencias e ingeniería. Esta biblioteca

se compone de diversas submódulos que abarcan áreas como optimización, álgebra lineal, integración, interpolación, procesamiento de señales, entre otros. SciPy es ampliamente utilizada por su versatilidad y potencia en el ámbito de la computación científica, y es considerada un aliado fundamental para matemáticos, científicos e ingenieros en el desarrollo de sus aplicaciones y proyectos (SciPy, 2023).

Pandas

Es una biblioteca de Python especializada en el manejo y análisis de datos. Ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales, lo que la hace ideal para tareas de limpieza, transformación y análisis de datos. Esta biblioteca es ampliamente utilizada en el ámbito de la ciencia de datos y la ingeniería, y proporciona herramientas poderosas y flexibles para trabajar con datos de manera eficiente y efectiva (Pandas, 2023).

NumPy

Es una biblioteca de Python que se utiliza para el cálculo numérico y el análisis de datos. Proporciona una estructura de datos llamada *array* que permite representar colecciones de datos y funciones muy eficientes para su manipulación. NumPy es ampliamente utilizado en el ámbito de la ciencia de datos y la ingeniería, y es considerado un componente fundamental en el desarrollo de aplicaciones y proyectos que involucran cálculo numérico y análisis de datos. Algunas de las funciones más populares de NumPy incluyen la creación de arrays, la manipulación de arrays, el álgebra lineal, la transformada de Fourier y la estadística básica (Harris et al., 2020).

3.2.3. Librerías para la Visualización de Datos

Matplotlib

Matplotlib es una biblioteca de Python que permite la creación de gráficos de línea de alto nivel y gráficos de píxel. Aunque es una biblioteca más baja en términos de nivel de abstracción en comparación con otras bibliotecas como Matplotlib, proporciona una interfaz de gráficos fácil de usar y flexible que admite una amplia gama de tipos de gráficos y estilos. Matplotlib es ampliamente utilizado en la comunidad de ciencia de datos y se integra fácilmente con otras bibliotecas como NumPy y SciPy.

Algunas características clave de Matplotlib incluyen:

Gráficos de línea: Matplotlib permite crear gráficos de línea unidimensional y bidimensional, lo que facilita la representación de datos numéricos y la identificación de tendencias.

Gráficos de píxel: Matplotlib también admite la creación de gráficos de píxel, que son útiles para visualizar imágenes digitales y datos espaciales.

Interoperabilidad: Matplotlib se integra fácilmente con otras bibliotecas de análisis de datos como NumPy y SciPy, lo que permite la creación de scripts de análisis de datos complejos y eficientes.

Personalización: Matplotlib ofrece una amplia gama de opciones de personalización que permiten a los usuarios ajustar el aspecto de los gráficos, como colores, estilos de líneas y símbolos de puntos.

A pesar de ser una biblioteca baja en términos de nivel de abstracción, Matplotlib proporciona una interfaz de gráficos fácil de usar y flexible que facilita la creación de gráficos de alta calidad y la visualización de datos (Matplotlib, 2023).

3.2.4. Librerías para Machine Learning

Scikit-Learn

Scikit-learn es una biblioteca de Python de aprendizaje automático de código abierto que proporciona herramientas para la construcción de modelos de aprendizaje automático. Scikit-learn se basa en NumPy y SciPy y proporciona una amplia gama de algoritmos de aprendizaje supervisado y no supervisado, así como herramientas para la selección de características, la validación cruzada y la visualización de resultados. Scikit-learn es ampliamente utilizado en el ámbito de la ciencia de datos y la ingeniería, y es considerado una de las bibliotecas de aprendizaje automático más populares y fáciles de usar en Python. Algunos de los algoritmos de aprendizaje automático más populares que se pueden implementar con Scikit-learn incluyen regresión lineal, clasificación, agrupamiento, reducción de dimensionalidad y análisis de componentes principales (PCA). Scikit-learn también proporciona herramientas para la evaluación y selección de modelos, lo que permite a los usuarios comparar y seleccionar el modelo más adecuado para sus datos (Scikit-Learn, 2023).

3.3. Análisis Exploratorio de Datos

El Exploratory Data Analysis (EDA) (*Exploratory Data Analysis*) es el proceso de análisis de un conjunto de datos para resumir sus principales características, a menudo utilizando métodos visuales. Puede realizarse para variables individuales, como trazar la distribución de una variable, o para relaciones entre variables, como crear un gráfico de dispersión (Hair et al., 2022). Este proceso no solo es esencial para comprender los datos, sino que también informa las decisiones sobre las técnicas estadísticas más adecuadas para el análisis posterior.

El proceso de EDA sigue una secuencia estructurada que comienza con los datos

Figura 3.3: Proceso de Desarrollo del Exploratory Data Analysis (EDA).



Fuente: Adaptado de *Aprende a realizar visualizaciones: Descubre la guía de introducción al análisis exploratorio* por Gobierno de España (2023).

brutos y avanza a través de varios pasos críticos antes de llegar a las conclusiones, como se muestra en la figura 3.3. Aquí está la descripción del proceso:

Datos Brutos: Se inicia con la recolección de datos en su estado más puro, sin procesar.

Análisis Descriptivo: Este es el primer paso en el EDA, donde se resumen las características principales de los datos. Se calculan estadísticas descriptivas como la media, mediana, moda, rango y desviación estándar. También se pueden crear visualizaciones iniciales, como histogramas y gráficos de barras, para entender la distribución de las variables individuales.

Ajuste de Tipos de Variables: Los datos se revisan para asegurar que cada variable tenga el tipo correcto (numérico, categórico, fecha/hora, etc.). Esto es crucial porque el tipo de variable determina qué tipo de análisis se puede realizar y qué herramientas estadísticas son apropiadas.

Detección y Tratamiento de Datos Ausentes: Se identifican los datos faltantes y se decide cómo manejarlos. Las opciones incluyen imputar valores, utilizar métodos para manejar datos ausentes o excluir observaciones o variables con demasiados valores faltantes.

Identificación de Datos Atípicos: Se detectan y examinan los valores atípicos, ya que pueden tener un impacto significativo en los resultados del análisis. Los datos atípicos pueden ser errores de entrada o variaciones naturales. La decisión de cómo tratarlos depende del contexto y del objetivo del análisis.

Correlación de Variables: Se analiza cómo las variables están relacionadas entre sí. Se utilizan técnicas como el coeficiente de correlación de Pearson para variables continuas o el coeficiente de correlación de Spearman para variables ordinales. Los gráficos de dispersión también son útiles para visualizar relaciones entre pares de variables.

Conclusiones: Los resultados y conocimientos obtenidos se comunican de manera efectiva a través de reportes, visualizaciones y presentaciones.

Cada uno de estos pasos es fundamental para garantizar que el análisis de datos sea robusto y confiable. El EDA es un proceso iterativo y a menudo revela información que guía la dirección del análisis posterior. Es una combinación de ciencia y arte que requiere curiosidad, escepticismo y un enfoque abierto para descubrir la verdad en los datos.

3.4. Algoritmos Machine Learning

En el Machine Learning, un algoritmo es básicamente un conjunto de instrucciones que un sistema de inteligencia artificial sigue para realizar tareas como identificar

patrones en los datos o hacer predicciones basadas en ciertas variables. Es la fórmula que define cómo se deben hacer los cálculos.

En contraste, un estimador es el resultado de aplicar ese algoritmo a un conjunto de datos específico. Es el modelo concreto que se ha entrenado y que se puede usar para hacer predicciones sobre nuevos datos. Así que mientras el algoritmo es el proceso general, el estimador es el producto final que se utiliza para inferir o predecir.

3.4.1. Algoritmo Decision Tree

Un árbol de decisiones (Decision Tree) en el aprendizaje automático es un modelo predictivo que mapea observaciones sobre un artículo a conclusiones sobre el objetivo del artículo. Es una estructura de árbol que se utiliza tanto para clasificación como para regresión.

Nodo Raíz: El proceso comienza con el nodo raíz, que contiene todo el conjunto de datos. Este nodo se divide en dos o más subconjuntos homogéneos basados en el atributo más significativo que se determina mediante medidas como la ganancia de información, el índice Gini o la reducción de la varianza.

División de Nodos: Cada nodo interno del árbol representa una “pregunta” o “condición” sobre alguna característica, y las ramas representan las posibles respuestas a esa pregunta. Este proceso de división se repite de forma recursiva en cada subconjunto derivado.

Nodos Hoja: Finalmente, los nodos terminales o nodos hoja representan la decisión o el resultado final. En un problema de clasificación, cada nodo hoja corresponderá a una etiqueta de clase, mientras que en un problema de regresión, representará

un valor continuo.

Construcción del Árbol: El árbol se construye de arriba hacia abajo de manera recursiva en un proceso conocido como partición binaria recursiva. Se selecciona la mejor división en cada paso hasta que se cumple un criterio de parada, como alcanzar una profundidad máxima del árbol o que no haya más ganancia en la mejora de las predicciones.

Poda del Árbol: Para evitar el sobreajuste, se puede realizar la poda del árbol, que elimina las partes del árbol que proporcionan poca potencia predictiva.

Tabla 3.1: Ventajas y Desventajas del Algoritmo DT

| | Ventajas | Desventajas |
|---|---|---|
| 1 | Fácil de entender e interpretar. | Puede sobreajustarse a los datos de entrenamiento. |
| 2 | No requiere mucha preparación de datos. | No es adecuado para relaciones complejas no lineales. |
| 3 | Puede manejar tanto datos numéricos como categóricos. | Sensible a variaciones en los datos de entrenamiento. |
| 4 | Utiliza un modelo de caja blanca interpretable. | La poda del árbol puede ser necesaria para evitar el sobreajuste. |
| 5 | Es posible validar un modelo utilizando pruebas estadísticas. | Puede ser menos preciso que otros modelos más complejos. |

Fuente: Elaboración Propia.

Los árboles de decisiones son populares debido a su facilidad de interpretación y comprensión. Son herramientas poderosas que pueden adaptarse a datos complejos y proporcionar una base para algoritmos más avanzados como los bosques aleatorios y el boosting.

3.4.2. Algoritmo Gradient Boosting Machine

El algoritmo de Gradient Boosting Machine (GBM) es una técnica poderosa de ML que se utiliza tanto para problemas de regresión como de clasificación. Funciona construyendo un modelo predictivo en forma de un conjunto de modelos de predicción débiles, que suelen ser árboles de decisión. A continuación, se describe el proceso de manera detallada:

Inicialización: GBM comienza con un modelo inicial, que puede ser tan simple como una predicción constante para todos los ejemplos de entrenamiento. Este modelo inicial se suele elegir para minimizar la función de pérdida que se está utilizando.

Construcción de Árboles Secuenciales: En cada iteración, GBM agrega un nuevo árbol de decisión que intenta corregir los errores cometidos por el modelo actual. Cada árbol se construye de manera que se enfoque en las áreas donde el modelo anterior tuvo un rendimiento deficiente.

Gradiente Negativo: Para cada árbol, se calcula el gradiente negativo de la función de pérdida con respecto a las predicciones del modelo actual. Esto indica en qué dirección se deben ajustar las predicciones para mejorar el modelo.

Aprendizaje de Árboles Débiles: Los árboles de decisión débiles, que suelen ser poco profundos, se ajustan a estos gradientes negativos. Cada árbol intenta predecir el gradiente negativo del modelo actual, actuando como un corrector de errores.

Reducción de la Tasa de Aprendizaje: GBM utiliza un parámetro llamado tasa de aprendizaje o "learning rate", que controla cuánto contribuye cada árbol al modelo

final. Una tasa de aprendizaje más baja puede hacer que el modelo sea más robusto y menos propenso al sobreajuste.

Adición de Árboles al Modelo: Después de aprender un árbol, se multiplica por la tasa de aprendizaje y se suma al modelo actual. Este proceso se repite iterativamente, agregando árboles uno por uno.

Regularización: Además de la tasa de aprendizaje, GBM puede utilizar otras técnicas de regularización, como la poda de árboles o la inclusión de solo un subconjunto de datos para entrenar cada árbol (potenciación estocástica), para evitar el sobreajuste.

Finalización: El proceso continúa hasta que se alcanza un número predefinido de árboles o hasta que la mejora en la función de pérdida es menor que un umbral establecido.

Tabla 3.2: Ventajas y Desventajas del Algoritmo GBM

| | Ventajas | Desventajas |
|---|--|--|
| 1 | Alta precisión predictiva. | Puede ser propenso al sobreajuste si no se regula adecuadamente. |
| 2 | Maneja bien los datos de diferentes tipos y distribuciones. | Requiere una selección cuidadosa de parámetros. |
| 3 | Puede capturar relaciones no lineales y complejas entre variables. | Más difícil de interpretar que los modelos más simples. |
| 4 | Robusto a los valores atípicos en el conjunto de datos de salida. | El tiempo de entrenamiento puede ser largo con grandes conjuntos de datos. |
| 5 | Mejora con el uso de técnicas de regularización y optimización. | Difícil de escalar con datos de alta dimensionalidad sin ajustes. |

Fuente: Elaboración Propia.

3.4.3. Algoritmo eXtreme Gradient Boosting

El algoritmo de eXtreme Gradient Boosting, conocido como **XGBoost**, es una técnica avanzada de ML que se basa en el principio de boosting.

Modelos Iniciales: xGBoost comienza construyendo un modelo simple, generalmente un árbol de decisión, y realiza predicciones sobre los datos de entrenamiento.

Cálculo de Errores: Después de cada predicción, xGBoost calcula los errores comparando las predicciones con los valores reales.

Construcción Secuencial: Se construye otro árbol para predecir y corregir estos errores. Este proceso se repite, con cada nuevo árbol construido para corregir los errores del anterior.

Optimización: Durante el entrenamiento, los parámetros de cada modelo se ajustan iterativamente utilizando un algoritmo de optimización como el descenso de gradiente, buscando minimizar una función objetivo, que puede ser el error de clasificación, el área bajo la curva (AUC), la raíz del error cuadrático medio (RMSE), entre otros.

Evaluación y Mejora: Cada modelo se compara con el anterior. Si un nuevo modelo tiene mejores resultados, se toma como base para realizar modificaciones. Si tiene peores resultados, se regresa al mejor modelo anterior y se modifica de una manera diferente.

Hiperparámetros: Los ajustes entre un modelo y otro son controlados por hiperparámetros que el usuario debe definir, como la tasa de aprendizaje.

Finalización: El proceso continúa hasta que la diferencia entre modelos consecutivos es insignificante, lo que indica que se ha encontrado el mejor modelo posible, o cuando se alcanza el número máximo de iteraciones definido por el usuario.

Tabla 3.3: Ventajas y Desventajas del Algoritmo xGBoost

| | Ventajas | Desventajas |
|---|--|--|
| 1 | Excelente rendimiento predictivo. | Puede ser propenso al sobreajuste sin la regularización adecuada. |
| 2 | Maneja eficientemente grandes conjuntos de datos. | Requiere una selección cuidadosa de hiperparámetros. |
| 3 | Puede manejar datos faltantes internamente. | Más complejo y difícil de interpretar que los modelos más simples. |
| 4 | Incluye regularización integrada para evitar el sobreajuste. | El tiempo de entrenamiento puede ser largo dependiendo de la complejidad del modelo. |
| 5 | Optimización basada en gradientes que mejora la precisión. | Puede ser desafiante para los principiantes debido a su complejidad. |

Fuente: Elaboración Propia.

xGBoost es conocido por su alta eficiencia, rendimiento y flexibilidad. Puede manejar problemas de regresión, clasificación binaria y multiclase, y admite funciones objetivas definidas por el usuario. Además, es altamente escalable, lo que lo hace adecuado para conjuntos de datos grandes y complejos.

3.4.4. Algoritmo Random Forest

El algoritmo Random Forest es una técnica de ML que opera creando un conjunto de árboles de decisión durante la fase de entrenamiento. Cada árbol se construye utilizando un subconjunto aleatorio de los datos y características, lo que ayuda a asegurar que los árboles sean variados y no correlacionados entre sí. El proceso

que realiza el algoritmo Random Forest es:

Selección de Subconjuntos: El algoritmo comienza seleccionando “n” números de registros aleatorios del conjunto de datos para cada árbol.

Construcción de Árboles: Se construyen árboles de decisión individuales para cada muestra seleccionada. Cada árbol es independiente y se basa en diferentes subconjuntos de datos y características.

Predicción de Cada Árbol: Cada árbol de decisión genera una salida o predicción basada en los datos que le fueron proporcionados.

Votación o Promedio: Para obtener la predicción final, el algoritmo combina las predicciones de todos los árboles. En tareas de clasificación, se utiliza la votación mayoritaria, y en tareas de regresión, se toma el promedio de las predicciones.

Resultado Final: La clase o valor que reciba la mayoría de los votos o el promedio se selecciona como la predicción final del modelo RF.

Los modelos ML como GBM, xGBoost y Random Forest tienen la ventaja de poder incorporar una amplia gama de variables operativas como la velocidad de rotación (RPM), el peso sobre la broca (WOB), la presión (PRESS), la profundidad del pozo (DEEP) y el torque (TQ). Estas variables son críticas para la optimización de la tasa de penetración (ROP) en tiempo real.

Tabla 3.4: Ventajas y Desventajas del Algoritmo RF

| | Ventajas | Desventajas |
|---|--|--|
| 1 | Puede manejar grandes conjuntos de datos con alta dimensionalidad. | Puede ser complejo y requerir mucho tiempo de computación. |
| 2 | Buen rendimiento y precisión en muchas tareas de predicción. | No es tan fácil de interpretar como un árbol de decisión individual. |
| 3 | Maneja bien los datos faltantes y no requiere escalamiento de características. | La gran cantidad de árboles puede hacer que el modelo sea difícil de implementar en tiempo real. |
| 4 | Reduce el riesgo de sobreajuste en comparación con los árboles de decisión individuales. | La creación de modelos puede ser menos intuitiva sin un conocimiento adecuado de los parámetros. |
| 5 | Proporciona estimaciones de la importancia de las características. | Puede ser menos efectivo en tareas donde las relaciones lineales son predominantes. |

Fuente: Elaboración Propia.

3.5. Métricas de Evaluación de un Modelo ML

Para evaluar un modelo de Machine Learning supervisado de regresión, te recomiendo utilizar las siguientes métricas, ya que cada una ofrece una perspectiva diferente sobre el rendimiento del modelo:

- Error Cuadrático Medio (MSE)
- Error Absoluto Medio (MAE)
- Raíz del Error Cuadrático Medio (RMSE)
- Coeficiente de Determinación (R^2)

3.5.1. Error Cuadrático Medio

El Error Cuadrático Medio (*Mean Squared Error (MSE)*) mide el promedio de los cuadrados de los errores, es decir, la diferencia cuadrática media entre los valores estimados y el valor real. Es muy sensible a los valores atípicos.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (3.1)$$

3.5.2. Error Absoluto Medio

El Error Absoluto Medio, en inglés *Mean Absolute Error (MAE)*, es el promedio de las diferencias absolutas entre los valores predichos y los observados. Proporciona una idea de la magnitud del error sin considerar su dirección.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (3.2)$$

3.5.3. Raíz del Error Cuadrático Medio

El Raíz del Error Cuadrático Medio (*Root Mean Squared Error (RMSE)*) es la raíz cuadrada del MSE y proporciona una medida de la magnitud del error en las mismas unidades que la variable de respuesta.

$$RMSE = \sqrt{MSE} \quad (3.3)$$

3.5.4. Coeficiente de Determinación R²

El Coeficiente de Determinación (R²) indica la proporción de la varianza de la variable dependiente que es predecible a partir de las variables independientes. Un valor

de R^2 cercano a 1 indica que el modelo explica una gran proporción de la varianza.

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (3.4)$$



Capítulo 4

MARCO APLICATIVO

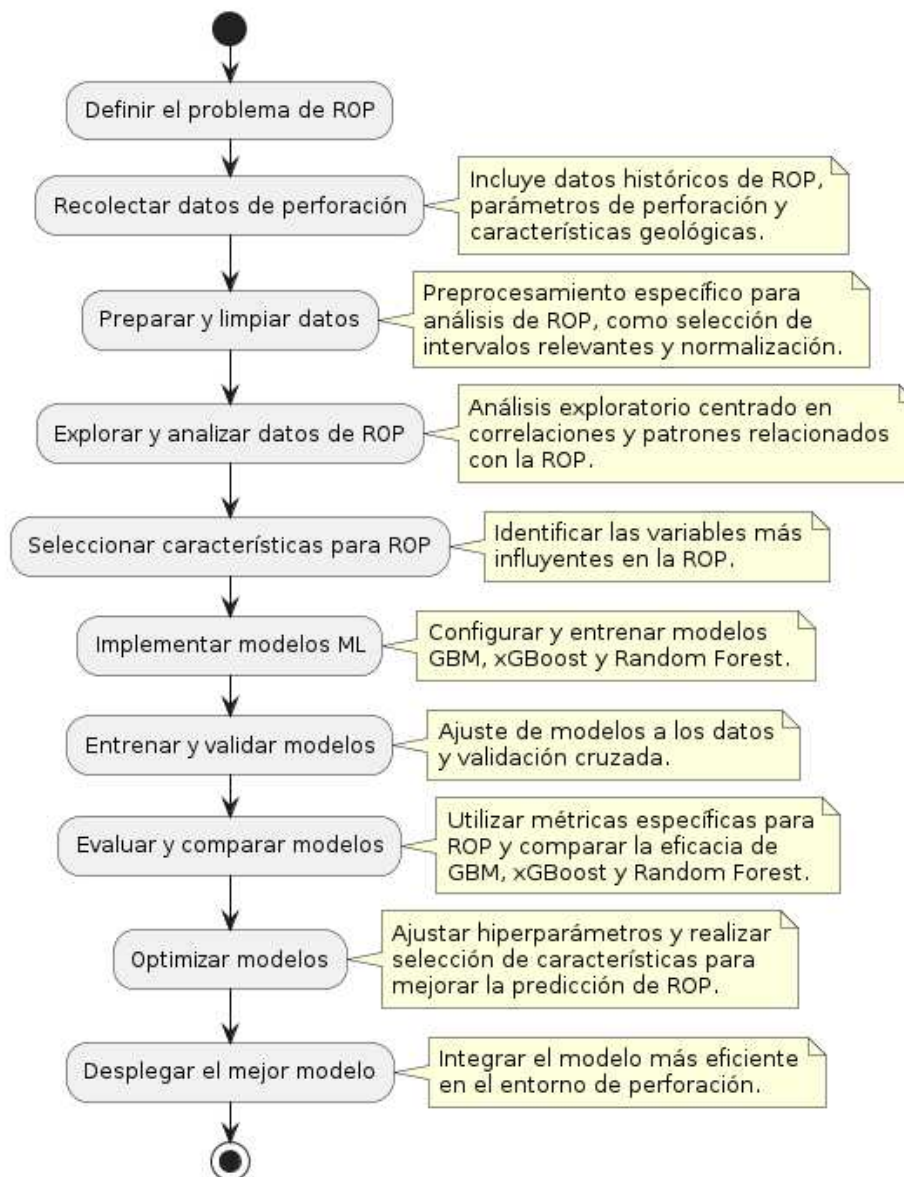
4.1. Introducción

El modelo ML será entrenado con los parámetros de perforación de los Pozos RGD-53, RGD-100D, SBL-8 y SRW-8D, con el cual se desarrollará los modelos ML.

El flujograma de la figura 4.1 para la optimización de la ROP describe un proceso sistemático y detallado que comienza con la definición del problema específico de ROP. Esto implica comprender los objetivos de optimización y las características únicas del entorno de perforación. A continuación, se recolectan datos de perforación relevantes, que incluyen información histórica, parámetros operativos y datos geológicos, que son esenciales para el análisis de ROP. Estos datos se preparan y limpian, lo que puede involucrar la selección de intervalos de datos pertinentes y la normalización para garantizar la calidad y la consistencia de los datos que se utilizarán en el modelado. En la segunda fase del flujograma, se lleva a cabo un análisis exploratorio para identificar patrones y correlaciones significativas relacionadas con la ROP. Basándose en este análisis, se seleccionan las características más influyentes que alimentarán los modelos de machine learning. Se implementan y entrenan

modelos avanzados como GBM, xGBoost y Random Forest, que son evaluados y comparados para determinar su eficacia en la predicción de ROP. El modelo con el mejor rendimiento se optimiza aún más mediante el ajuste de hiperparámetros y la selección de características.

Figura 4.1: Proceso de Análisis de Datos y Desarrollo de Modelo ML para Optimización de ROP.



Fuente: Elaboración Propia.

4.2. Descripción Geológica

4.2.1. Campo San Antonio

El Bloque San Antonio se encuentra ubicado entre las Serranías Agüaragüe y Caipipendi de la faja Subandina Sur. El campo de Sábalo ubicado en el anticlinal del mismo nombre es parte de la serranía de San Antonio, formado en el ambiente compresivo de la Faja plegada y corrida del Sub Andino Sur Boliviano.

Estratigrafía

Las unidades estratigráficas que están involucradas en el campo Sábalo del Bloque San Antonio, abarcan desde el Silúrico hasta el Terciario. Las formaciones Huamampampa, Icla y Santa Rosa (Devónico), constituyen los reservorios, compuestos por intercalaciones de lutitas y areniscas de baja porosidad primaria, depositadas en sistemas de lóbulos deltaicos, y de cara de playa predominantemente.

Las formaciones Huanampampa y Santa Rosa son cuerpos arenosos importantes y distribuidos en gran parte de la cuenca. Por el contrario, la formación Icla es predominantemente arcillosa y en su parte media presenta un cuerpo arenoso, la cual se encuentra conectada hidráulicamente con Huamampampa.

La disposición de los perfiles de pozo permite identificar dos grupos de paquetes estratigráficos de acuerdo a su composición litológica. Estos tipos son:

- Paquetes espesos y masivos de areniscas.
- paquetes con una alternancia de areniscas y lutitas.

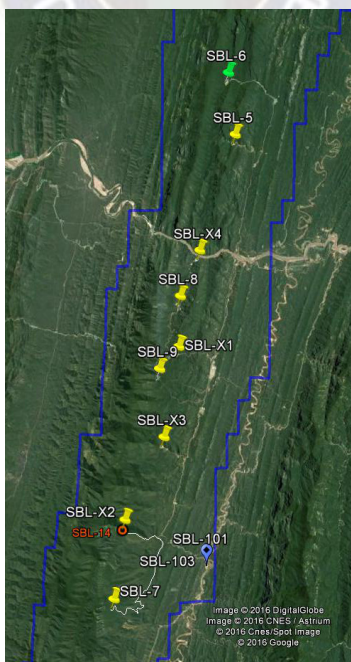
Los paquetes masivos son regionalmente constantes y los cambios fisiológicos y de espesor por lo común son transicionales. Los paquetes con alternancia de areniscas y lutitas están integrados por una sucesión de parasecuencias de progradación.

Cada una de estas parasecuencias presenta una secuencia de progradación bien desarrollada y un abrupto episodio transgresivo. Eventos de regresiones forzadas, aunque no son lo común, también están registrados dentro las secuencias atravesadas por los pozos.

Ubicación de Pozos

El campo Sábalo tiene más de dieciocho años continuos de producción y cuenta a la fecha con 10 pozos productores de gas activos (incluyendo el multilateral SBL-7), los cuales se encuentran en producción continua. Durante el 2012 se perforó el pozo inyector SBL-101i para descarte de agua y durante el 2014 el pozo SBL-103i, iniciando la inyección en la Formación Tapehua a partir de septiembre de 2014. En abril de 2016 dió inicio la perforación del pozo SBL-6 ubicado al norte del campo e inicio producción 29 de abril del 2018. En la figura 4.2 se muestra la ubicación de los pozos del campo Sábalo, incluyendo el pozo SBL-8 que es de nuestro interés.

Figura 4.2: Ubicación de Pozos del Campo San Antonio.



Fuente: Agencia Nacional de Hidrocarburos (ANH).

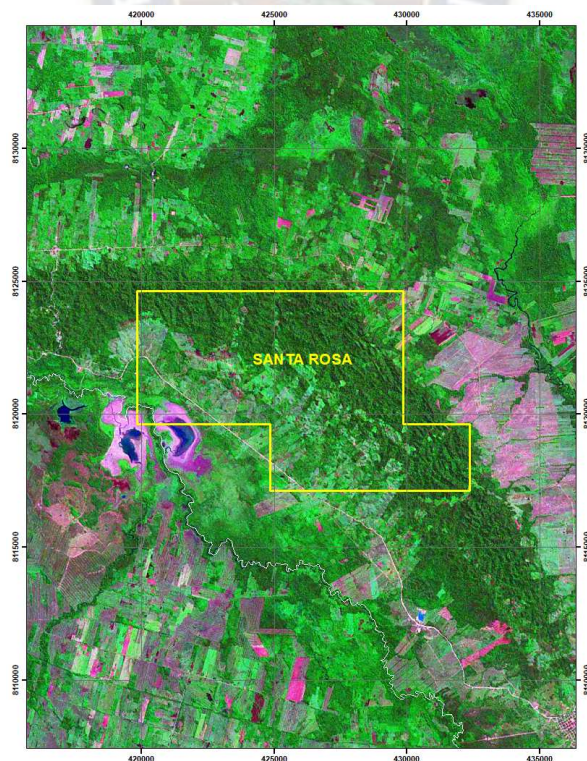
4.2.2. Campo Santa Rosa

El campo Santa Rosa de la figura 4.3, está situado en Santa Cruz, fue descubierto por YPFB en 1982 con el pozo SRW-X6. Los pozos SRW-X4 y SRW-X5 se añadieron después, siendo el SRW-X5 productor de gas y condensado en las formaciones Ayacucho y Arenisca, mientras que los otros dos produjeron solo en Ayacucho. La rehabilitación de los tres pozos fue exitosa.

Ubicación de Pozos

El campo Santa Rosa tiene como productores a los reservorios Arenisca 1, Arenisca 2, Ayacucho. Los reservorios Piray y Sara cesaron producción por alto corte de agua. Actualmente cuenta con 6 pozos productores. El campo inició producción en junio de 2009.

Figura 4.3: Campo Santa Rosa.



Fuente: Agencia Nacional de Hidrocarburos (ANH).

4.2.3. Campo Río Grande

El campo Río Grande mostrado en la figura 4.4, se encuentra al sur este de la ciudad de Santa Cruz de la Sierra. Se encuentra ubicado dentro del Área de Contrato homónimo, cerca al límite de deformación influenciada por la tectónica del subandino. Este campo cuenta con varios reservorios ubicados en niveles someros del Cretácico (Cajones) y otros más profundos del Carbonífero (San Telmo, Escarpment) y devónico (Iquiri), todos productores de Gas y Condensado.

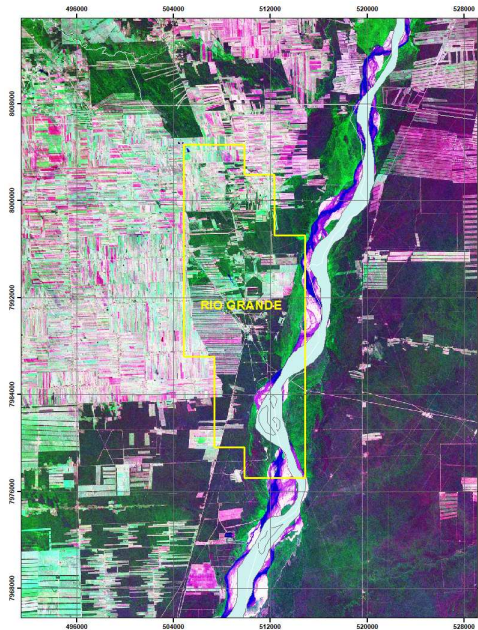
Estratigrafía

La estratigrafía de la zona de Río Grande es conocida, basada en la información de los pozos perforados en el área. Las formaciones Terciarias identificadas son: Guandacay, Tariquia, Yecua y Petaca. Las formaciones Cretácicas son: Cajones, Yantata e Ichoa. La única formación del Pérmico es Elvira. Las formaciones del Carbonífero son: San Telmo, Escarpment, Taiguati, Chorro, Tarija y Tupambi y la formación devónica alcanzada fue Iquiri.

Ubicación de Pozos

En total se perforaron 90 pozos perforados, de los cuales la mayor parte se concluyó con arreglos dobles: 23 pozos son productores, 36 cerrados, dos inyectores de agua, tres sin terminación y 26 abandonados. Se inyecta gas al reservorio Taiguati “Y” procedente del mismo campo y de Los Sauces. El campo Río Grande tiene los siguientes reservorios: Petaca Medio, Petaca Inferior, Taiguati O Superior Norte, Taiguati O Inferior Norte, Taiguati W, Taiguati X, Taiguati Y, Tupambi Sud y Tupambi Norte.

Figura 4.4: Campo Rio Grande.



Fuente: Agencia Nacional de Hidrocarburos (ANH).

4.3. Análisis Exploratorios de Datos

4.3.1. Análisis de Datos

Para el desarrollo de un modelo ML primeramente se optó realizar el análisis de datos con la mayor cantidad de datos posibles, para ello se utiliza los datos de parámetros de perforación de los pozos: RGD-100D, RGD-53, SBL-8 y SWR-8D con el objetivo de obtener un modelo único. Posteriormente se observa la calidad de los datos como ser: datos anómalos, datos faltantes, distribución de los datos entre otros, de esta forma elegir los hiperparámetros para el modelo con el objetivo de reducir su varianza y sesgo respecto a los datos reales.

Entre los parámetros de perforación registrados y recopilados de diferentes pozos se tiene un total de 7926 datos con 13 variables como se puede observar a continuación:

- PROF (m) profundidad del pozo

- ROP (min/m) tasa de penetración
- ROP (m/hr) tasa de penetración
- WOB (klb) peso sobre el trépano
- RPM (rpm) revolución por minuto
- FLOW IN (gpm) flujo de entrada de lodo
- PRESS (psi) presión
- TQ (lbs/ft) torque
- DENS IN (lpg) densidad de entrada de lodo
- DENS OUT (lpg) densidad de salida de lodo
- ECD (lps) densidad equivalente de circulación
- GAS (ugt) unidad de gas total
- POZO

Para el análisis de datos se utilizan las librerías de Python: matplotlib, numpy, pandas, seaborn, scikit-learn y xgboost.

```
1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5
```

Listing 4.1: Importación de Librerías

En la tabla 4.1 se puede observar los datos registrados:

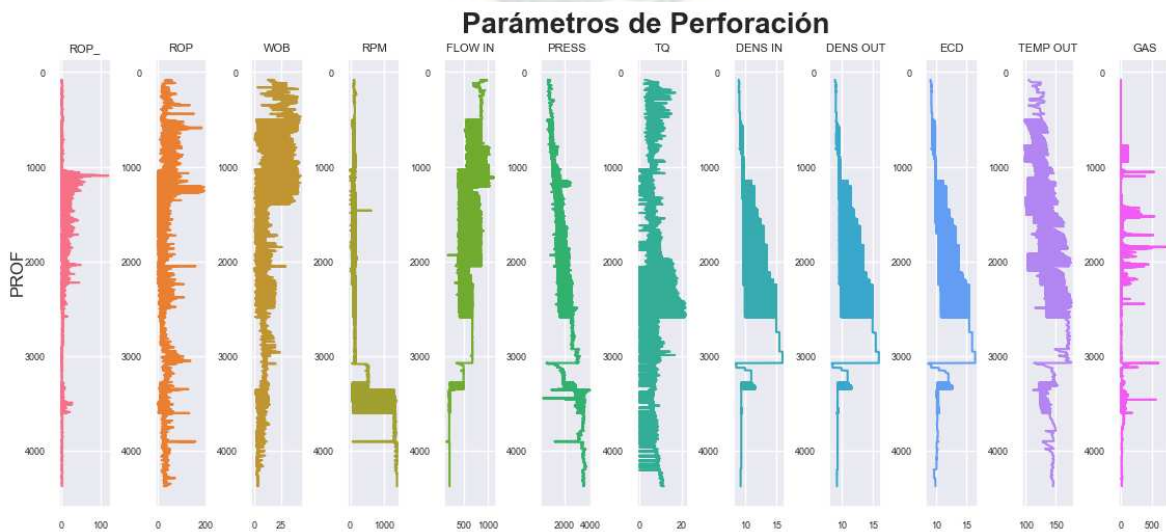
Tabla 4.1: Registro de Parámetros de Perforación

| | PROF | ROP_ | ROP | WOB | RPM | FLOW IN | PRESS | TQ | DENS IN | DENS OUT | ECD | TEMP OUT | GAS | POZO |
|------|------|-------|-------|------|--------|---------|--------|------|---------|----------|-------|----------|--------|----------|
| 1 | 509 | 1.3 | 46.15 | 3 | 61.2 | 547.1 | 634.38 | 5.39 | 9.2 | 9.2 | 9.6 | 98.16 | 0 | RGD-100D |
| 2 | 510 | 1.67 | 36.02 | 2 | 60.2 | 547.1 | 649.04 | 5.29 | 9.2 | 9.2 | 9.6 | 98.17 | 0 | RGD-100D |
| 3 | 511 | 0.9 | 66.67 | 4 | 60 | 547.1 | 685.27 | 5.37 | 9.2 | 9.2 | 9.6 | 98.19 | 0 | RGD-100D |
| 4 | 1523 | 15.79 | 3.8 | 3.2 | 64 | 402 | 1834 | 7 | 9.7 | 9.7 | 9.9 | 114 | 489.67 | SRW-8D |
| 5 | 1524 | 17.14 | 3.5 | 4 | 64 | 402 | 1838 | 6.2 | 9.7 | 9.7 | 9.9 | 114 | 389.65 | SRW-8D |
| 6 | 1525 | 15.79 | 3.8 | 3.5 | 64 | 402 | 1826 | 4.2 | 9.7 | 9.7 | 9.9 | 114 | 454.69 | SRW-8D |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 7921 | 3581 | 4.4 | 13.64 | 9.7 | 80 | 214 | 3057 | 6.4 | 9.5 | 9.5 | 10.4 | 127 | 7.99 | RGD-53 |
| 7922 | 3582 | 8.5 | 7.06 | 10 | 80 | 214 | 3067 | 6.5 | 9.5 | 9.5 | 10.4 | 127 | 5.06 | RGD-53 |
| 7923 | 3583 | 9.4 | 6.38 | 10 | 80 | 216 | 3040 | 6.2 | 9.5 | 9.5 | 10.4 | 127 | 6.2 | RGD-53 |
| 7924 | 2542 | 1.85 | 32.4 | 9.8 | 113.05 | 681 | 2550 | 0 | 15 | 15 | 15.41 | 171 | 3.78 | SBL-8 |
| 7925 | 2543 | 2.03 | 29.5 | 11.4 | 112.88 | 680 | 2560 | 0 | 15 | 15 | 15.41 | 170.5 | 3.2 | SBL-8 |
| 7926 | 2544 | 3.66 | 16.4 | 8.5 | 153.21 | 682 | 2630 | 6.6 | 15 | 15 | 15.41 | 170.2 | 2.66 | SBL-8 |

Fuente: Agencia Nacional de Hidrocarburos (ANH).

El conjunto de datos se visualiza en una gráfica de registro en la figura 4.5 donde podemos observar cómo estos parámetros de perforación cambian con la profundidad.

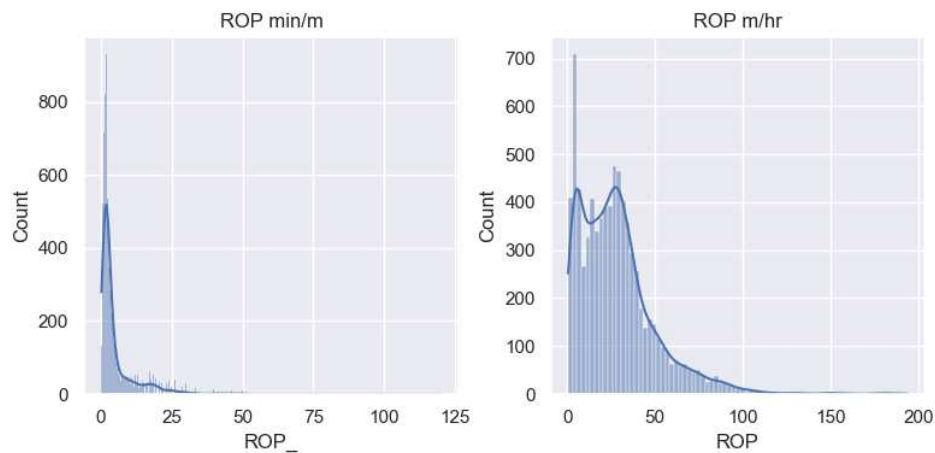
Figura 4.5: Parámetros de Perforación de Pozos



Fuente: Agencia Nacional de Hidrocarburos (ANH)

Dado que se tienen dos registros de Rate of Penetration (ROP) pero en diferentes unidades se procede a analizar con mayor detalle esta característica, ya que es nuestro parámetro objetivo. En la figura 4.6 se puede observar que la ROP (min/m) posee mayor distribución, por lo que para el análisis de la correlación con los demás parámetros se tomará en cuenta este registro.

Figura 4.6: Histograma de la Rate of Penetration (ROP)



Fuente: Elaboración Propia.

Mediante un gráfico de dispersión mostrada en la figura 4.7 podemos observar que:

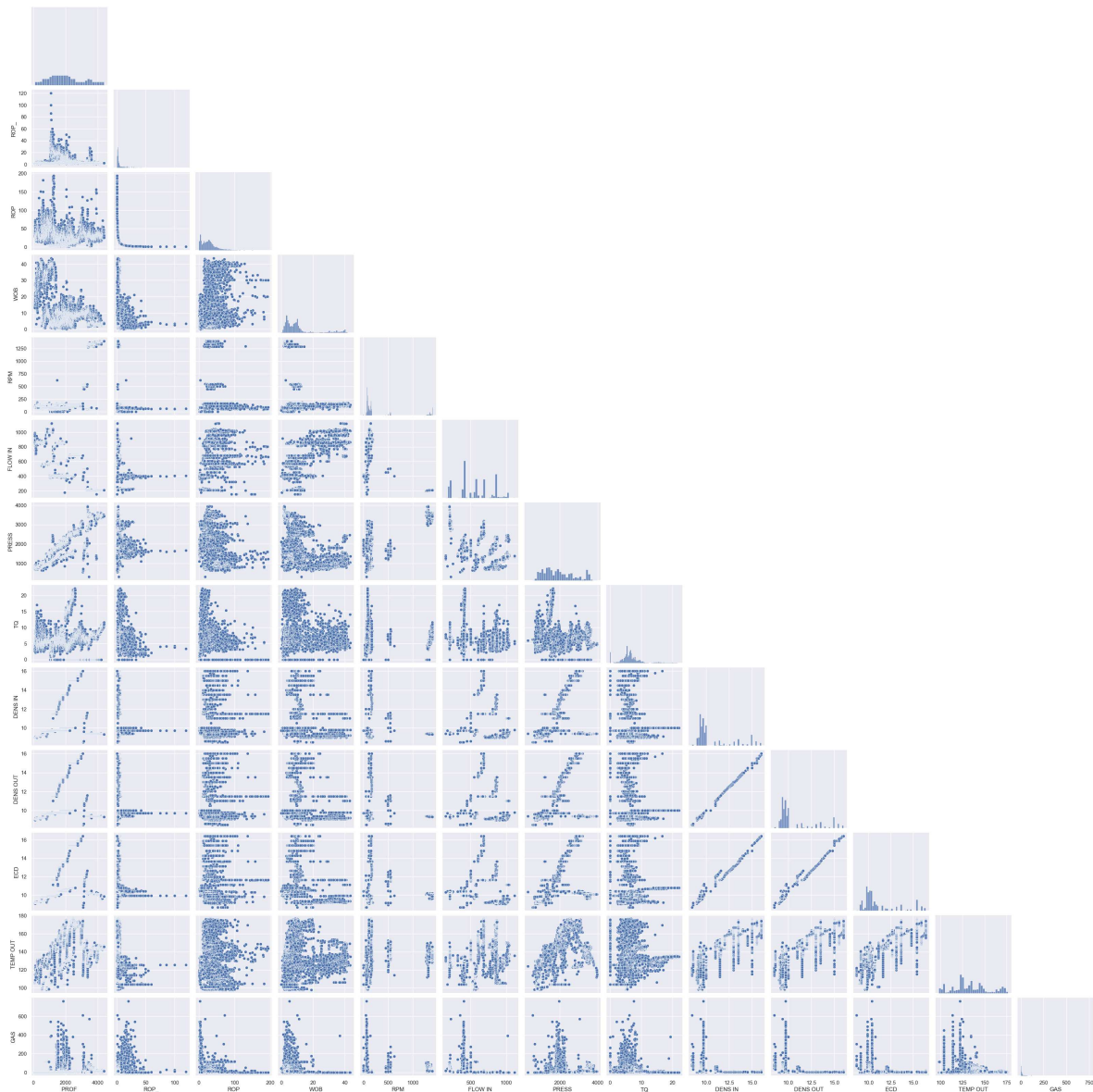
Correlación Visual: Observamos una dispersión de puntos en el gráfico. La relación entre las variables no parece ser lineal, lo que sugiere que un modelo más complejo podría ser necesario. No hay una tendencia clara en la dirección vertical (ROP) en función de la dirección horizontal (otra variable).

Outliers: Algunos puntos están alejados de la nube principal. Estos podrían ser valores atípicos o datos anómalos que afectan la predicción.

Homogeneidad de la Dispersión: La dispersión de puntos parece ser más densa en ciertas áreas. Esto podría indicar agrupamientos o patrones específicos.

Selección de Características: Para predecir la ROP, debemos considerar qué otras variables están relacionadas con ella. Las variables con una correlación más fuerte podrían ser candidatas importantes para incluir en el modelo.

Figura 4.7: Diagrama de dispersión de los parámetros de perforación



Fuente: Elaboración Propia.

Un mapa de calor es una representación gráfica de datos en la que se utiliza el color para mostrar la densidad o intensidad de un fenómeno en una superficie bidimensional. Esta técnica resalta áreas con altas o bajas concentraciones de

información, lo que facilita la identificación de patrones, tendencias y anomalías en los datos.

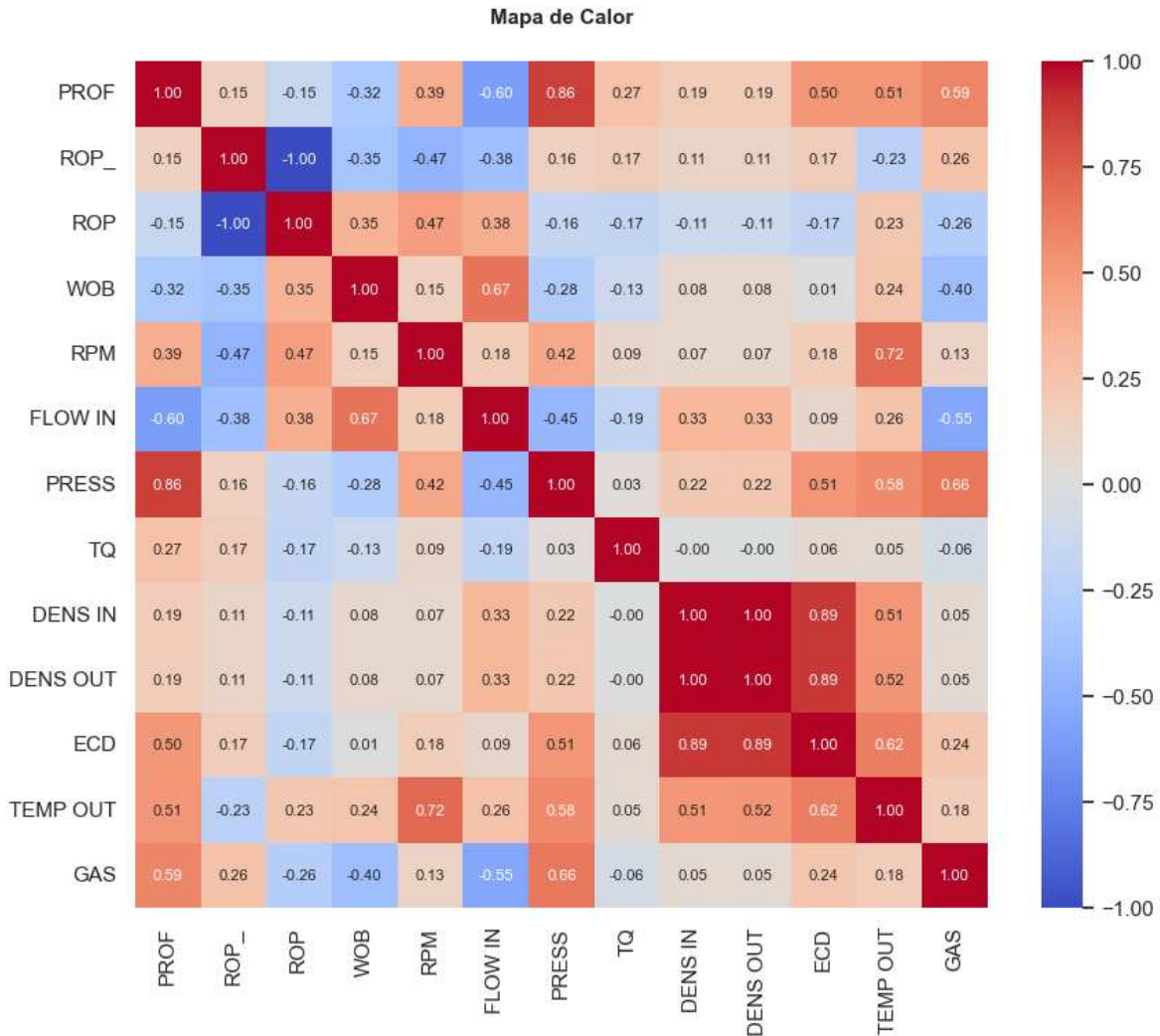
Es necesario señalar que el Mapa de Calor de la figura 4.8 y el diagrama de Dispersión de la figura 4.7 ofrecen informaciones similares por lo que la diferencia destaca, sobretodo en nuestro caso es muy notable, en la facilidad de lectura; en el mapa de calor podemos observar que variables están relacionadas, en cambio en el diagrama de Dispersión se puede observar como se relacionan estas variables (lineal, exponencial, etc). Por otra parte en el Mapa de Calor, un color mucho más pronunciado significa una mayor relación de variables ya sea negativa (azul) o positiva (rojo), sin embargo cuando una variable se relaciona consigo misma no aporta ninguna información y esto se observa en la diagonal de la figura 4.8.

Usando el método de la correlación de Spearman, y mostrando en un mapa de calor como en la figura 4.8, en la cual solo se analiza la fila o columna perteneciente a la ROP ya que es nuestra variable objetivo y solo las variables que poseen mayor relación con la ROP se tomara en cuenta.

En es sentido, se puede observar que el parámetro WOB posee mayor relación sobre los otros parámetros, así mismo los parámetros GAS, PRESS y FLOW IN no poseen información relevante para la ROP por lo que se descarta como una variable a ser analizado. Por otra parte en un diagrama de dispersión de los parámetros no es clara una tendencia apreciable respecto a la ROP, al contrario se observa la existencia de datos atípicos, por lo que se procederá a analizarlo con mayor detalle.

Usando el diagrama de cajas como en la figura 4.9 nuevamente se observa la irrelevancia de los parámetros GAS, PRESS, FLOW IN, además de los datos atípicos por lo que no se tomará en cuenta en el desarrollo del modelo ML ya que podrían proveer información anómala.

Figura 4.8: Correlación de los Parámetros de Perforación

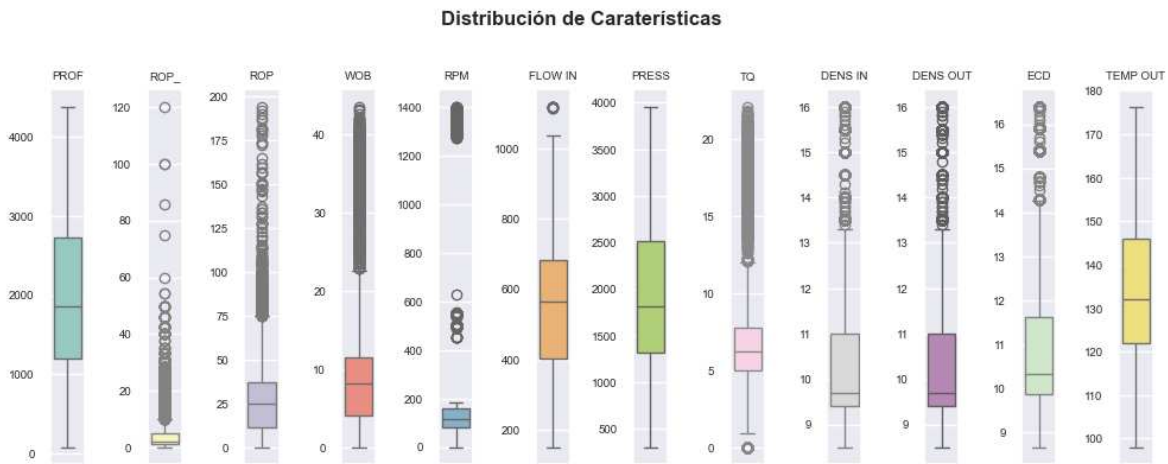


Fuente: Elaboración Propia.

Finalmente realizamos una descripción rápida de los datos que tenemos luego del análisis exploratorio de datos teniendo solo 7620 datos de las cuales las mas relevantes son PROF, ROP, WOB, TQ, ECD, así mismo realizamos una nueva correlación con el método de Spearman, de la cual se observa que la ROP está mas fuertemente relacionado con la WOB y ECD, esto quiere decir que existe una relación que no necesariamente es lineal.

El mapa de calor de correlación de Spearman mostrada en la figura 4.8 proporciona información valiosa sobre las relaciones entre diferentes parámetros. Analizando

Figura 4.9: Diagrama de cajas de los Parámetros de Perforación



Fuente: Elaboración Propia.

el mapa de calor proporcionado, aquí hay algunas observaciones clave sobre la relación entre las variables y la ROP.

WOB y PRESS: Estas variables tienen una correlación positiva moderada con la ROP (WOB 0.47, PRESS 0.44). Esto sugiere que un aumento en el peso sobre la broca (WOB) o la presión (PRESS) podría resultar en un aumento en la tasa de penetración (ROP).

RPM y FLOW IN: Ambas muestran una correlación negativa moderada con la ROP (RPM -0.35, FLOW IN -0.38). Esto indica que un incremento en las revoluciones por minuto (RPM) o el flujo de entrada (FLOW IN) podría asociarse con una disminución en la ROP.

Correlaciones Bajas o Insignificantes: Las demás variables presentan correlaciones bajas con la ROP, lo que implica que no tienen una influencia tan directa o significativa sobre la ROP como las variables mencionadas anteriormente.

Estas relaciones son importantes para considerar al desarrollar un modelo de machine

learning para predecir la ROP, ya que indican qué variables podrían ser predictores significativos y cómo podrían influir en la predicción.

4.4. Modelo predictivo Machine Learning

Para el modelo predictivo se hará uso de los algoritmos Gradient Boosting Machine (GBM), eXtreme Gradient Boosting (xGBoost) y Random Forest (RF), además se usará la librería de scikit-learn y xGBoost.

4.4.1. Modelo predictivo con el algoritmo GBM

El puntaje del Score de entrenamiento fue del 89% y el puntaje del Score de prueba fue del 80%.

```
1 # Separate feature and target
2 from sklearn.model_selection import train_test_split
3 from sklearn.preprocessing import StandardScaler
4 from sklearn.pipeline import Pipeline
5 from sklearn.ensemble import GradientBoostingRegressor
6
7 # Separate feature and target
8 X = df.drop(['ROP'], axis=1)
9 y = df['ROP']
10 # Train test split
11 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
12 # Make pipeline
13 steps = [('scaler', StandardScaler()),
14          ('gbr', GradientBoostingRegressor(min_samples_leaf=4, max_depth=10))]
15 model_GBM = Pipeline(steps)
16
17 # Fit pipeline to training data
18 model_GBM.fit(X_train, y_train)
```



```

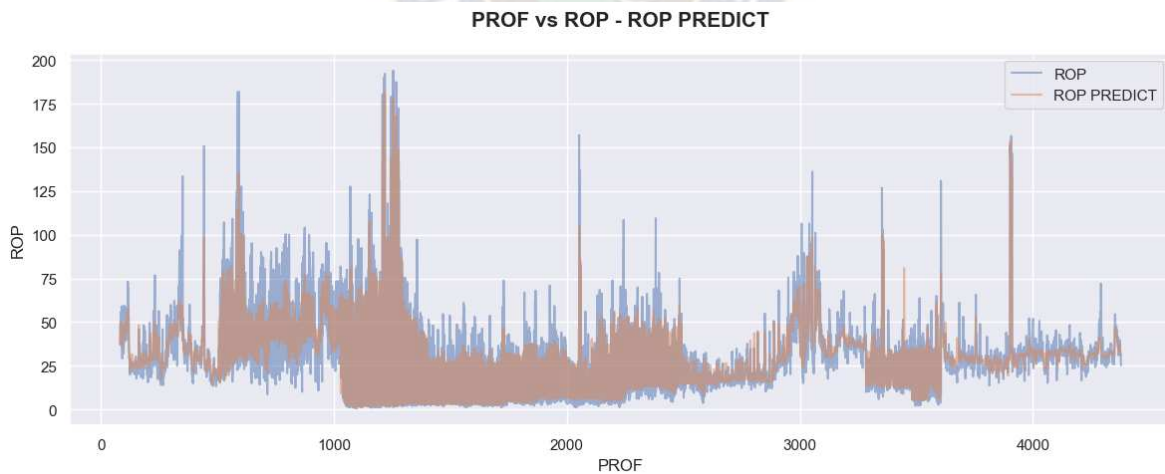
19
20 # Evaluate model with R2 metric
21 train_score = model_GBM.score(X_train, y_train)
22 test_score = model_GBM.score(X_test, y_test)
23 print(f'Score on train set: {train_score:.2f}')
24 print(f'Score on test set: {test_score:.2f}')
25

```

En la figura 4.10 que compara la Rate of Penetration (ROP) registrada con la ROP predecida por un modelo de machine learning GBM, donde se puede observar una correlación entre los dos conjuntos de datos. Sin embargo, no es perfecta. Ambos tienen picos prominentes, pero no siempre coinciden en magnitud o posición exacta.

El modelo GBM tiene cierta capacidad para prever la ROP, pero aún puede mejorarse para aumentar su precisión y confiabilidad.

Figura 4.10: Comparación de la ROP registrado y la ROP predecido con el modelo GBM



Fuente: Elaboración Propia.

4.4.2. Modelo predictivo con el algoritmo xGBoost

El puntaje del Score de entrenamiento fue del 86 % y el puntaje del Score de prueba fue del 74 %.

```
1  from xgboost import XGBRegressor
2  from sklearn.model_selection import train_test_split, GridSearchCV
3  from sklearn.metrics import mean_squared_error
4
5  X = df.drop(['ROP'], axis=1)
6  y = df['ROP']
7  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
8
9  model_xgb = XGBRegressor(random_state=42)
10 parameters_xgb = {
11     'n_estimators': [100, 500, 1000],
12     'max_depth': [2, 4, 6],
13     'colsample_bytree': [0.2, 0.6, 0.8],
14     'min_child_weight': [3, 5, 7],
15     'gamma': [0.3, 0.5, 0.7],
16     'subsample': [0.4, 0.6, 0.8]
17 }
18
19 gridsearch_xgb = GridSearchCV(estimator=model_xgb, param_grid=parameters_xgb)
20 gridsearch_xgb.fit(X_train, y_train)
21
22 print(f'Mejores hiperparametros: {gridsearch_xgb.best_params_}')
23
24 best_model_xgb = gridsearch_xgb.best_estimator_
25 predicciones = best_model_xgb.predict(X_test)
26
27 train_score_xgb = best_model_xgb.score(X_train, y_train)
28 test_score_xgb = best_model_xgb.score(X_test, y_test)
29
```

```

30 print('Score on train: ', train_score_xgb)
31 print('Score on test: ', test_score_xgb)
32

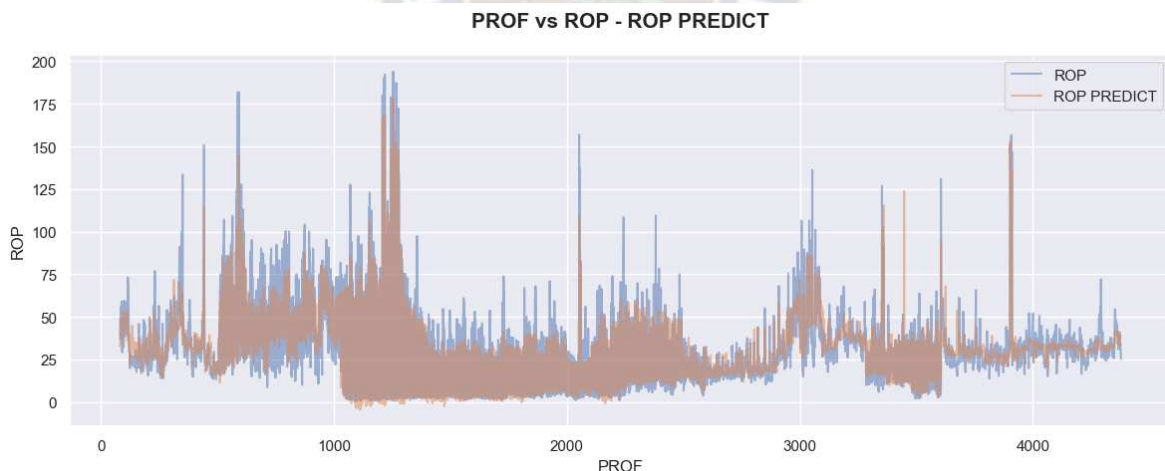
```

En la figura 4.11 que compara la Rate of Penetration (ROP) registrada con la ROP predecida por un modelo de xGBoost, donde se puede observar:

Correlación Visible: Se observa una correlación entre los dos conjuntos de datos tal como sucede con el modelo GBM.

Precisión del Modelo xGBoost: El modelo xGBoost tiene cierta capacidad para predecir la ROP con una precisión de 74 %, se redujo ligeramente la varianza sin embargo el sesgo en comparación al modelo GBM.

Figura 4.11: Comparación de la ROP registrado y la ROP predecido con el modelo xGBoost



Fuente: Elaboración Propia.

4.4.3. Modelo Predictivo con el Algoritmo Random Forest

El puntaje del Score de entrenamiento fue del 97 % y el puntaje del Score de prueba fue del 80 %.

```

1  from sklearn.ensemble import RandomForestRegressor
2  from sklearn.model_selection import train_test_split, GridSearchCV
3  from sklearn.metrics import mean_squared_error
4
5  X = df.drop(['ROP'], axis=1)
6  y = df['ROP']
7
8  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
9  model_rf = RandomForestRegressor(random_state=42)
10
11  parameters_rf = {
12      'n_estimators': [100, 200, 300],
13      'max_depth': [None, 10, 20, 30],
14      'min_samples_split': [2, 5, 10],
15      'min_samples_leaf': [1, 2, 4],
16      'max_features': ['auto', 'sqrt', 'log2']
17  }
18
19  gridsearch_rf = GridSearchCV(estimator=model_rf, param_grid=parameters_rf)
20  gridsearch_rf.fit(X_train, y_train)
21  print(f'Mejores hiperparametros: {gridsearch_rf.best_params_}')
22
23  best_model_rf = gridsearch_rf.best_estimator_
24  predicciones = best_model_rf.predict(X_test)
25
26  train_score_rf = best_model_rf.score(X_train, y_train)
27  test_score_rf = best_model_rf.score(X_test, y_test)
28
29  print('Score on train: ', train_score_rf)
30  print('Score on test: ', test_score_rf)
31

```

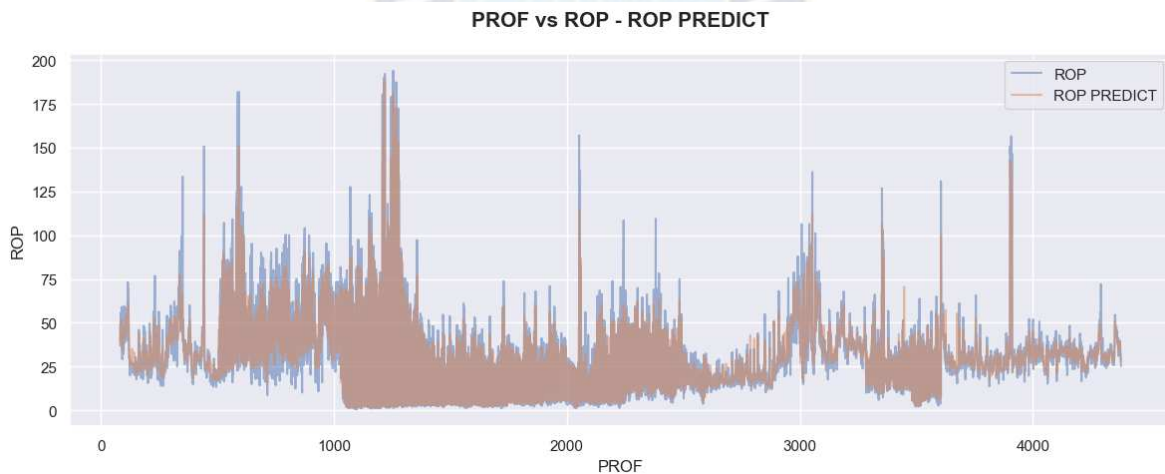
En la figura 4.12 que compara la ROP registrada con la ROP predecida por un

modelo de RF, donde se puede observar:

Correlación Visible: Sugiere que no existe una diferencia significativa respecto a los anteriores modelos a primera vista, sin embargo Random Forest sería un modelo candidato junto con el modelo xGBoost, cuando se trata de mayor precisión y esto seguramente se reflejará ante grandes volúmenes de datos.

Precisión del Modelo Random Forest: El modelo Random Forest obtuvo un Score de prueba de 80 %, ligeramente mejor que el modelo GBM pero cuando se trata de varianza sin duda el modelo xGBoost es el primer candidato.

Figura 4.12: Comparación de la ROP registrado y la ROP predicho con el modelo Random Forest



Fuente: Elaboración Propia.

A pesar de la correlación, se deben considerar factores operativos, como condiciones del pozo, cambios en la formación y otros eventos imprevistos. La tabla 4.2 muestra la comparación del Score de los modelos ML, en cual xGBoost se posiciona mejor obteniendo una menor varianzarespecto a los demás, sin embargo tratandose de precisión Random Forest es el mejor posicionado.

La tabla 4.3 se muestra una comparaciónde las métricas de evaluación de los diferentes modelos. De acuerdo a $MSE = 104.761$ y $MAE = 6.111$ el modelo Random Forest

Tabla 4.2: Tabla de Comparación de los Modelos ML

| | Train Score | Test Score |
|-------------------------------------|--------------------|-------------------|
| Gradient Boosting Machine (GBM) | 0.89 | 0.80 |
| eXtreme Gradient Boosting (xGBoost) | 0.84 | 0.76 |
| Random Forest | 0.97 | 0.80 |

Fuente: Elaboración Propia.

está mejor posicionado indicando que es el menos sensible a los errores grandes. Así mismo mediante el RMSE = 10.235 de obtenida por Random Forest es el que menor error posee y mejor se asemeja a los datos reales con un $R^2 = 0.807$, en resumen Random Forest es el que menor sesgo obtuvo.

Tabla 4.3: Tabla de Métricas de los Modelos ML

| | MSE | MAE | RMSE | R2 |
|-------------------------------------|------------|------------|-------------|-----------|
| Gradient Boosting Machine (GBM) | 109.869 | 6.494 | 10.482 | 0.798 |
| eXtreme Gradient Boosting (xGBoost) | 137.274 | 7.261 | 11.716 | 0.747 |
| Random Forest | 104.761 | 6.111 | 10.235 | 0.807 |

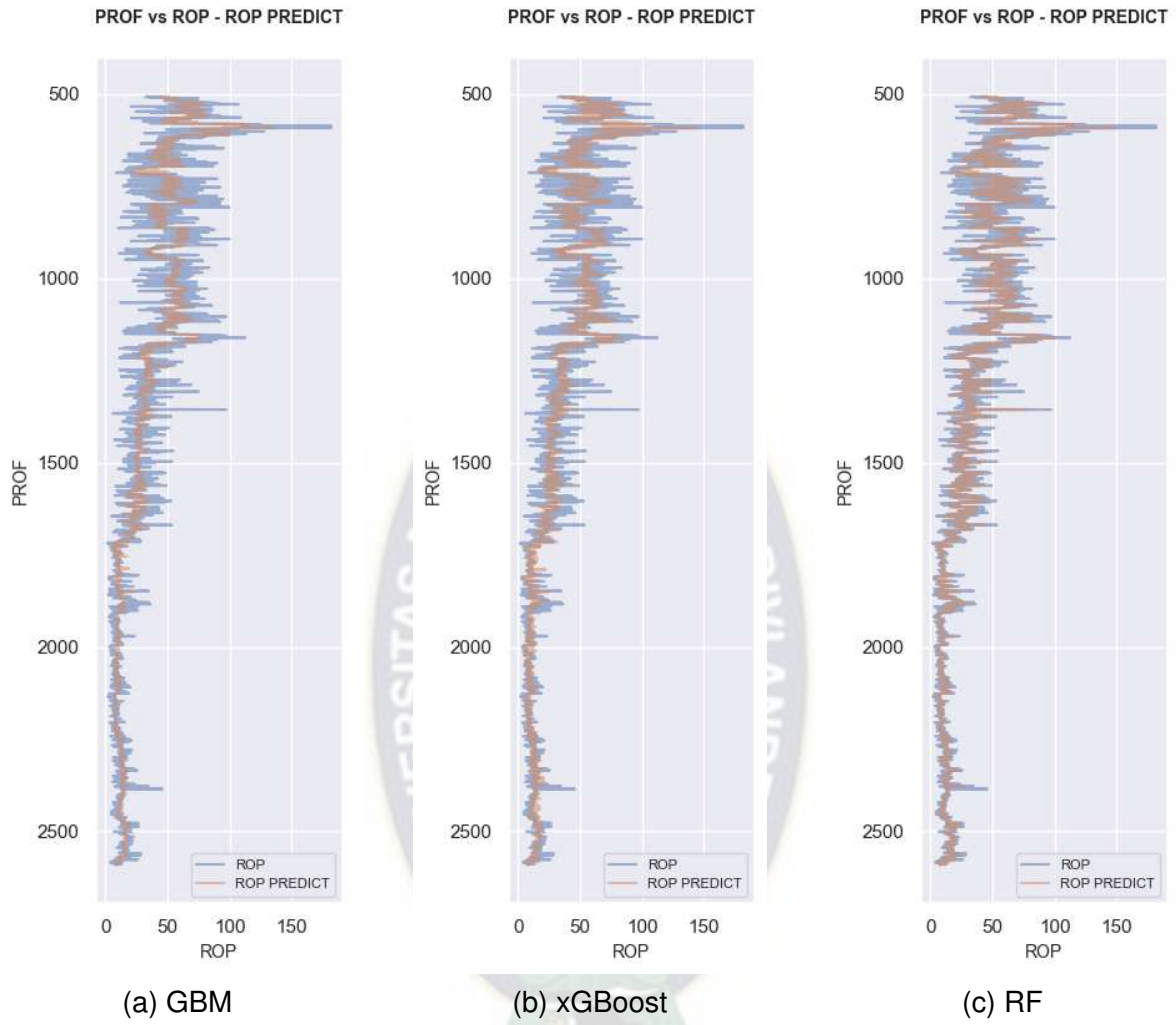
Fuente: Elaboración Propia.

4.5. Aplicación de los Modelos en el Pozo RGD-100D

La figura 4.13 que muestra una comparación de la ROP del Pozo RGD-100D entre los valores reales y predichos utilizando tres modelos diferentes: GBM, xGBoost y Random Forest.

La figura 4.14 que muestra una comparación del exponente Dc de perforación del Pozo RGD-100D entre los valores reales y predichos utilizando tres modelos

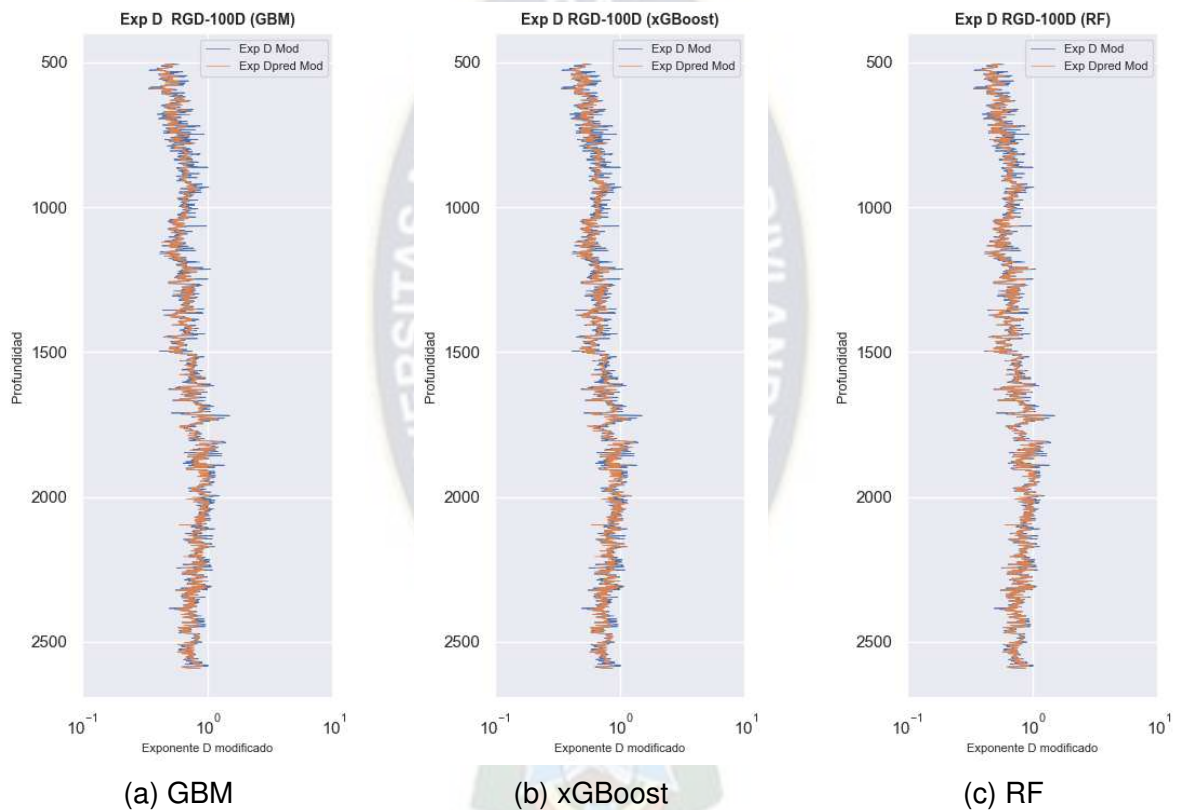
Figura 4.13: Gráfica de la ROP del Pozo RGD-100D aplicando los modelos GBM 4.13a, xGBoost 4.13b y RF 4.13c.



Fuente: Elaboración Propia.

diferentes: GBM, xGBoost y Random Forest.

Figura 4.14: Gráfica del Exponente Dc del Pozo RGD-100D aplicando los modelos GBM 4.14a, xGBoost 4.14b y RF 4.14c.



Fuente: Elaboración Propia. Adaptado de «Aplicación del Método Machine Learning en la Toma de Decisiones para Optimizar la Perforación de Pozos» por Condori (2024).

Capítulo 5

ANÁLISIS DE COSTOS

5.1. Estimación de Costo

Implementar soluciones de AI personalizadas puede variar entre 6000 \$us hasta los 300000 \$us, dependiendo de varios factores como se muestra en la tabla 5.1 como el tipo de IA, la complejidad de la solución y el volumen de datos. Mientras que las soluciones estándar son más económicas, las personalizadas, aunque más costosas, ofrecen una adaptabilidad total a las necesidades del negocio (Codster, 2024).

En el contexto de la industria petrolera, el costo de desarrollar un modelo de machine learning puede ser muy variable. Factores como la precisión requerida y la cantidad y complejidad de los datos pueden influir significativamente en el costo. Por ejemplo, soluciones que requieren alta precisión, como un sistema de reconocimiento facial con un 99% de fiabilidad, tendrán un costo mucho mayor. Por lo que el costo de desarrollar un modelo de machine learning es variable y depende directamente del alcance y las especificaciones del proyecto, la tecnología a usarse, facilidades técnicas, así como del sector industrial en el que se implemente, como la industria

Tabla 5.1: Estimación de Costos para el Desarrollo e Implementación del Modelo de ML en un proyecto mediano.

| Concepto | Descripción | Costo (USD) |
|--------------------------|-------------------------------------|----------------|
| Investigación y Diseño | Desarrollo inicial del modelo | 20,000 |
| Adquisición de Datos | Recopilación de datos necesarios | 10,000 |
| Herramientas y Software | Licencias y herramientas | 5,000 |
| Salarios de Expertos | Científicos de datos e ingenieros | 100,000 |
| Capacitación | Formación en nuevas tecnologías | 5,000 |
| Integración de Sistemas | Incorporación del modelo a sistemas | 15,000 |
| Pruebas y Validación | Verificación del modelo | 10,000 |
| Mantenimiento del Modelo | Mantenimiento anual | 10,000 |
| Soporte Técnico | Asistencia técnica anual | 10,000 |
| Administración | Gestión del proyecto | 10,000 |
| Infraestructura | Servidores y almacenamiento | 20,000 |
| Reserva para Imprevistos | Contingencias | 20,500 |
| Total | | 225,500 |

Fuente: Codster.

petrolera. Para obtener una estimación más precisa, es recomendable consultar con especialistas y considerar las necesidades específicas del proyecto que muchas veces son a largo plazo.

5.2. Ventajas y Desventajas

Ventajas:

- Mejora de la Eficiencia: La aplicación de *Machine Learning* puede aumentar significativamente la eficiencia de la perforación al predecir y ajustar las operaciones

en tiempo real.

- Reducción de Costos: Al optimizar el proceso, se pueden reducir los costos operativos y de inversión al evitar perforaciones innecesarias o incorrectas.
- Toma de Decisiones Basada en Datos: El *Machine Learning* proporciona una base sólida para la toma de decisiones, utilizando grandes volúmenes de datos para hacer predicciones precisas.

Desventajas:

- Complejidad Tecnológica: Requiere conocimientos especializados en ML y perforación de pozos, lo que puede ser un desafío en términos de capacitación y desarrollo de habilidades.
- Dependencia de Datos: La calidad y cantidad de datos históricos disponibles pueden limitar la precisión del modelo.
- Inversión Inicial: Puede haber una inversión inicial significativa en tecnología y recursos humanos para desarrollar e implementar el modelo.
- Resistencia al Cambio: Puede haber resistencia por parte de los trabajadores y empresas del sector debido a la transición a nuevas tecnologías y métodos de trabajo.

El análisis de costos puede proporcionar una visión integral de los gastos requeridos para el desarrollo, implementación y mantenimiento de un proyecto. En el caso de nuestro trabajo de mediana inversión, que implica el desarrollo de un modelo de ML para la optimización de la perforación de pozos, el análisis de costos nos permite estimar la inversión necesaria y planificar adecuadamente para la asignación de recursos.

5.3. Tiempo No Productivo NPT

El Tiempo No Productivo (*del inglés Non-Productive Time (NPT)*), es el tiempo que se pierde en la operación de perforación debido a situaciones que no estaban contempladas en el programa original. Las causas del NPT son diversas y pueden incluir:

- Fallas de equipos: Roturas de herramientas, problemas con la maquinaria o fallos en los sistemas de control.
- Condiciones geológicas: Dificultades para perforar formaciones duras, presencia de fluidos no deseados o inestabilidad del pozo.
- Eventos imprevistos: Accidentes, condiciones climáticas adversas o problemas de seguridad.

El NPT tiene un impacto negativo significativo en los proyectos de perforación, ya que genera retrasos, aumenta los costos y reduce la rentabilidad (Ayala et al., 2017). Por lo tanto, la minimización del NPT es un objetivo primordial para las empresas petroleras, por lo que la optimización de la ROP mediante la AI busca reducir el NPT y por ende los costos asociados.

5.4. Análisis de Costos

El Campo Sábalo, ubicado en el bloque San Antonio en Bolivia, es uno de los campos de gas natural más importantes del país. Su desarrollo ha sido un proceso gradual que se ha llevado a cabo a lo largo de varias décadas, con la participación de diferentes empresas petroleras. En la tabla 5.2 se tiene el costo de algunos pozos que se encuentran en el Campo Sábalo, donde se aprecia que los costos varían de acuerdo a la profundidad perforada y que influye en el tiempo de perforación de manera

proporcional, pero en promedio se tiene que el costo de perforación varía de entre 6000 \$us a 8000 \$us.

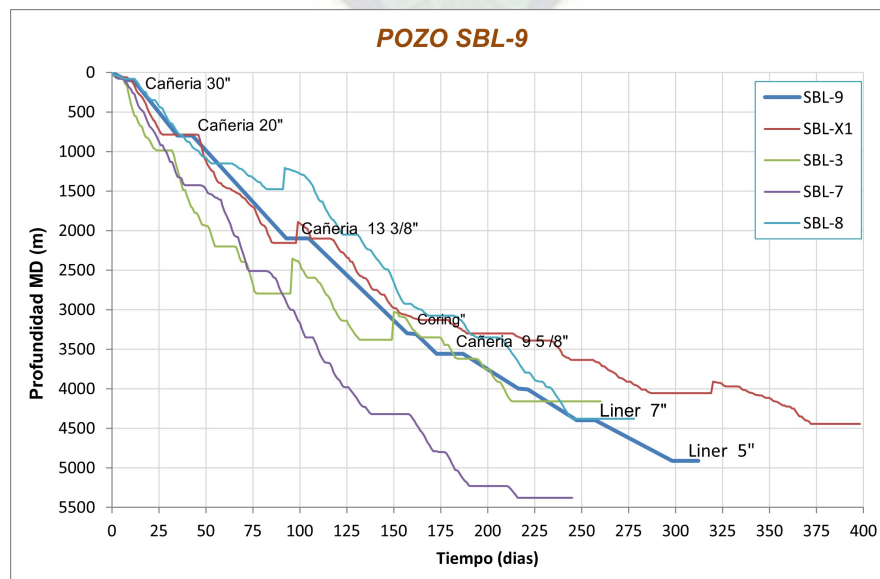
Tabla 5.2: Costo de Perforación de los pozos en el Campo Sábalo

| Pozo | Profundidad (m) | Costo de Perforación \$us | Dias de Perforación |
|--------|-----------------|---------------------------|---------------------|
| SBL-X1 | 4443 | 34429634 | 397 |
| SBL-X2 | 5264 | 32542048 | 447 |
| SBL-X3 | 4346 | 23424424 | 133 |
| SBL-7 | 5380 | 43916515 | 245 |
| SBL-8 | 4378 | | 278 |
| SBL-9 | 4935 | 48248310 | 206 |

Fuente: Petrobras Bolivia.

Independientemente del pozo, siempre existirá el NPT que pueden estar vinculados a diferentes factores imprevistos, eso se observar en la Curva de Avance de los diferentes pozos Campo Sábalo en la figura 5.1, y fue el pozo SBL-X1 el más afectado con el NPT teniendo un tiempo de duración de 397 días, al contrario de la curva de avance del pozo SBL-7 que no tuvo NPT considerables.

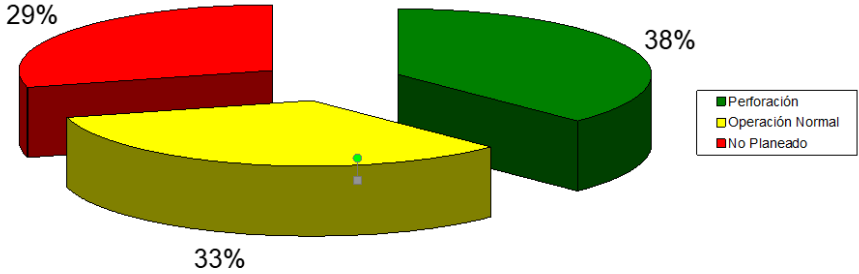
Figura 5.1: Curva de Avance de los pozos en el Campo Sábalo



Fuente: Agencia Nacional de Hidrocarburos (ANH) YPFB Andina.

Asi mismo se muestra en la figura 5.2 la distribución de tiempo de perforación en el Campo Sábalo donde se puede observar que el 29% del tiempo de perforación se centra en la NPT y como consecuencia tiene un costo aproximado de al menos 28746000 \$us solo entre los pozos SBL-X1, SBL-x2, SBL-X3, SBL-7, SBL-8 y SBL-9.

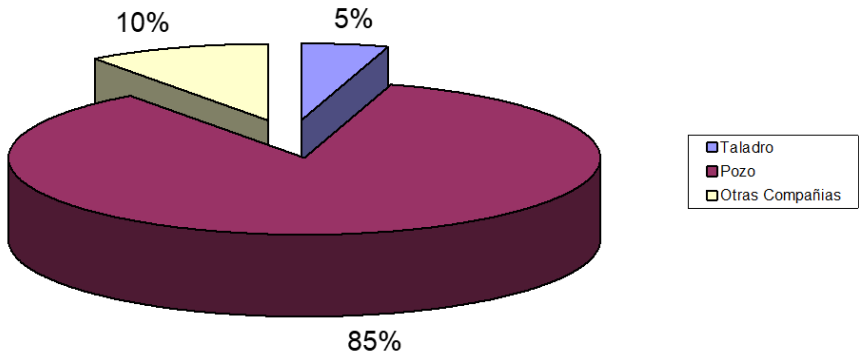
Figura 5.2: Distribución de tiempos de perforación en el Campo Sábalo



Fuente: Petrobras Bolivia.

Es necesario señalar que no todo NPT está asociado a ROP como se muestra en la figura 5.3, el 5% corresponde al trépano, 10% a otras compañías y el 85% corresponde a NPT asociado al pozo que es lo que se busca reducir con la optimización la ROP y se traduce en un costo de 24434100 \$us solo entre los pozos SBL-X1, SBL-x2, SBL-X3, SBL-7, SBL-8 y SBL-9.

Figura 5.3: Distribución de NPT en el Campo Sábalo



Fuente: Petrobras Bolivia.

Capítulo 6

CONCLUSIONES Y RECOMENDACIONES

6.1. Conclusiones

- Se ha logrado una comprensión de las variables críticas que influyen en la eficiencia del avance de perforación. Como se pudo observar en la figura 4.8 los valores WOB y RPM entre los más correlacionados con la ROP, esto dió lugar a la realización de un análisis de los datos adquiridos como son los parámetros de perforación de pozos.
- La investigación ha demostrado que la tecnología de ML es una herramienta valiosa en el ámbito de la perforación de pozos. La implementación de prácticas de ML no solo ha permitido identificar patrones y tendencias en los datos de perforación, sino que también ha facilitado la adopción de prácticas óptimas que pueden ser replicadas para mejorar los procesos así también permitió seleccionar los modelos ML que mejor se adaptan de acuerdo a los datos adquiridos. Este trabajo en particular se optó por modelos de regresión como

son GBM, xGBoost y Random Forest, con los cuales se ha desarrollado y validado un modelo predictivo de ML utilizando un conjunto de datos representativos obteniendo un Score promedio de 80 % pero no se puede pasar por desapercibido que aun existe una notable varianza de aproximadamente 10 %. Estos modelos han demostrado ser capaz de predecir con precisión el avance de la perforación a partir de una cantidad de datos y variables reducidas, proporcionando una herramienta crucial para la toma de decisiones en tiempo real siendo este la ventaja frente a los modelos tradicionales y por ende la mejora continua de las operaciones de perforación.

- Los modelos GBM, xGBoost y RF han presentado resultados similares en términos de la ROP y la gráfica del exponente D sin embargo en términos de despliegue de Software, los modelos xGBoost y Random Forest son los que mejores resultados obtuvieron por lo tanto son modelos a hacer énfasis de estudio. Por otra parte el exponente D_c expresa una tendencia positiva durante una perforación, en este caso en pozo RGD-100D, lo cual indica una perforación normal. Un cambio de tendencia del exponente D sería indicativo de un evento anómalo durante la perforación y los modelos obtenidos son capaces de detectar este evento, sus causas requerirá otro tipo de análisis como por ejemplo las propiedades petrofísicas de las formaciones.

La consistencia en el rendimiento refleja la robustez de los modelos de ML en la predicción de parámetros críticos de perforación y subraya su potencial para ser aplicados en diferentes escenarios de perforación.

Este estudio ha aportado al ámbito de la perforación de pozos, ofreciendo perspectivas diferentes y valiosas para obtener herramientas predictivas que pueden emplearse para optimizar las operaciones y promover una mayor eficiencia y efectividad en la industria. El trabajo ha resultado ser positivo, especialmente en la adquisición

de conocimientos, la comprensión de variables y la aplicación de técnicas de ML respecto a la perforación de pozos. No obstante, para una evaluación final completa, es esencial considerar otros factores como la aplicabilidad práctica de los resultados y su influencia en la eficiencia y seguridad de las operaciones de perforación.

6.2. Recomendaciones

Si bien se logró un Score destacable de los modelos en la aplicación de pozos, el estudio de las formaciones podría ser significativo para mejorar y/o ajustar el modelo, así mismo el estudio y evaluación de algoritmos como Redes Neuronales, Soporte de Vectores o incluso el Aprendizaje Profundo podrían proporcionar una mejor perspectiva en las predicciones de los parámetros de perforación así como también las causas específicas de los eventos anormales. Por otra parte el Análisis de Datos es crucial para el desarrollo de un modelo Machine Learning (ML), por lo que una gran cantidad de datos (datos completos de un pozo y/o campo) y su respectivo estudio es indispensable para una mejor interpretación de los parámetros de perforación.

Bibliografía

- Adams, N., & Charrier, T. (1985). *Drilling engineering: a complete well planning approach*.
- AWS, A. W. S. (2022). *What is Reinforcement Learning? - Reinforcement Learning Explained - AWS* [Última actualización 2022-07-18 UTC]. <https://aws.amazon.com/what-is/reinforcement-learning/>
- Ayala, D., Torres, H., Valencia, R., & Loaiza, M. (2017, enero). *Impacto del Tiempo no Productivo en operaciones de perforación y análisis de los datos mediante la prueba de Chicuadrado* (Vol. 14). <https://doi.org/10.18273/revfue.v14n2-2016001>
- Azar, J., & Samuel, G. (2007). *Drilling Engineering*. PennWell Corporation. <https://books.google.com.bo/books?id=eseViO982VgC>
- Belyadi, H., & Haghghat, A. (2021). *Machine Learning Guide for Oil and Gas Using Python: A Step-by-Step Breakdown with Data, Algorithms, Codes, and Applications*. Elsevier Science. <https://books.google.es/books?id=MjoEEAAQBAJ>
- Bingham, M. (1965). *A New Approach to Interpreting Rock Drillability*. Petroleum Publishing Company. <https://books.google.com.bo/books?id=8KcRnQEACAAJ>
- Brenjkar, E., & Biniáz Delijani, E. (2022). *Computational prediction of the drilling rate of penetration (ROP): A comparison of various machine learning approaches and traditional models* (Vol. 210). Journal of Petroleum Science; Engineering. <https://doi.org/https://doi.org/10.1016/j.petrol.2021.110033>

- Chilingar, G., Serebryakov, V., & Robertson, J. (2002). *Origin and Prediction of Abnormal Formation Pressures*. Elsevier. <https://books.google.com/books?id=naVTAAAAMAAJ>
- Codster. (2024). *¿Es caro desarrollar una inteligencia artificial?* <https://codster.io/blog/es-caro-desarrollar-una-inteligencia-artificial/>
- Condori, M. L. E. (2024). Aplicación del Método Machine Learning en la Toma de Decisiones para Optimizar la Perforación de Pozos [Código Fuente del Proyecto]. https://github.com/LuisCM17/ml_ropPrediction
- DataTab. (2024). *Tutorial sobre gráficos*. <https://datatab.es/tutorial/charts>
- Education, I. C. (2020, julio). *Machine Learning*. IBM.
- EnergyAPI. (2001). *Manual de Fluidos de Perforación*.
- Farnós, J. D. (2023, septiembre). *El aprendizaje profundo y el machine learning generan escenarios de transformación educativos facilitados por procesos algorítmicos y neuronales artificiales*. Consultado el 14 de enero de 2024, desde <https://juandomingofarnos.wordpress.com/2023/09/14/el-aprendizaje-profundo-y-el-machine-learning-generan-escenarios-de-transformacion-educativos-facilitados-por-procesos-algoritmicos-y-neuronales-artificiales/>
- Gil, L. A. (2020). *Aprende Machine Learning con Python: Aprendizaje Supervisado*. Independently Published.
- Gobierno de España. (2023). *Aprende a realizar visualizaciones: Descubre la guía de introducción al análisis exploratorio*. datos.gob.es. Consultado el 14 de enero de 2024, desde <https://datos.gob.es/es/boletines/aprende-realizar-visualizaciones-descubre-la-guia-de-introduccion-al-analisis-exploratorio>
- Hair, J., Babin, B., Anderson, R., & Black, W. (2022). *Multivariate Data Analysis*. Cengage Learning.
- Hammack, R. L., Cooper, R. S., Hruday, S. E., Kamp, K. E., & Sharp, J. G. (2010). *Applied Petroleum Reservoir Engineering*.

- Harris, C., Millman, K., van der Walt, S., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M., Brett, M., Haldane, A., del Río, J., Wiebe, M., Peterson, P., ... Oliphant, T. (2020). *Array programming with NumPy* (Vol. 585). Nature. <https://doi.org/10.1038/s41586-020-2649-2>
- Hawker, D. (2001). *Análisis de las presiones anormales de formación* (Vol. 2). Datalog.
- Hawker, D., VOGT, K., & Robinson, A. (2001). *Manual de Operaciones en el Pozo: Procedimientos y operaciones en el pozo* (Vol. 3). Datalog.
- IBM. (2024). *Unsupervised Learning*. <https://www.ibm.com/es-es/topics/unsupervised-learning>
- Jeffery, W. (2015). *Deep Well Drilling: The Principles and Practices of Deep Well Drilling and a Hand Book of Useful Information for the Well Driller (Classic Reprint)*. FB&C Limited.
- Matplotlib, D. T. (2023). *Matplotlib* [Última actualización 2023-12-08 UTC]. <https://matplotlib.org/stable/index.html>
- Microsoft. (2024). *Microsoft impulsa la transformación digital de la minería, petróleo y gas en Bolivia* [Accedido el 21 de marzo de 2024]. Consultado el 21 de marzo de 2024, desde <https://news.microsoft.com/es-xl/microsoft-impulsa-la-transformacion-digital-de-la-mineria-petroleo-y-gas-en-bolivia/>
- Mitchell, R., Miska, S., Aadnøy, B., & of Petroleum Engineers (U.S.), S. (2011). *Fundamentals of Drilling Engineering*. Society of Petroleum Engineers. <https://books.google.com.bo/books?id=v1keKQEACAAJ>
- Oracle. (2023). *¿Qué es la IA?* <https://www.oracle.com/es/artificial-intelligence/what-is-ai/>
- Pandas, D. T. (2023). *Pandas Documentation* [Última actualización 2023-12-08 UTC]. <https://pandas.pydata.org/docs/>

- Scikit-Learn, D. T. (2023). *User Guide: Contents — Scikit-Learn 1.3.2 Documentation* [Última actualización 2023-12-08 UTC]. https://scikit-learn.org/stable/user_guide.html
- SciPy, D. T. (2023). *User Guide: Contents — SciPy v1.11.4 Reference Guide* [Última actualización 2023-11-18 UTC]. <https://docs.scipy.org/doc/scipy/tutorial/index.html#user-guide>
- Serrano, A. (2016). *Inteligencia Artificial: Fundamentos, Prácticas y Aplicaciones*. Alpha Editorial. <https://books.google.com.bo/books?id=03J6EAAAQBAJ>
- Short, J. (1993). *Introduction to Directional and Horizontal Drilling*. PennWell Books.
- Sobhi, I., Dobbi, A., & Hachana, O. (2022). *Prediction and analysis of penetration rate in drilling operation using deterministic and metaheuristic optimization methods* (Vol. 12). Springer.
- White, F. M. (2016). *Fluid Mechanics*. McGraw-Hill Education.

Apéndice A

Fundamentos de Estadística y Análisis de Datos con Python

A.1. Introducción a la Estadística

A.1.1. Definición y Importancia

La estadística es una rama de las matemáticas que se ocupa de la recolección, análisis, interpretación y presentación de masas de datos numéricos. Es fundamental para la toma de decisiones basadas en datos y para la comprensión de fenómenos complejos a través de la simplificación numérica.

Población Conjunto completo de elementos o eventos que se están estudiando.

En estadística, se representa con el símbolo:

$$N \tag{A.1}$$

Muestra Subconjunto seleccionado de la población. Se representa con el símbolo:

$$n \quad (A.2)$$

A.1.2. Tipos de Datos Estadísticos

Datos Cualitativos Describen cualidades o características que no pueden ser medidas con números.

Datos Cuantitativos Se refieren a cantidades numéricas que se pueden medir.

A.1.3. Fórmulas Básicas en Estadística

Media Aritmética

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (A.3)$$

Donde \bar{x} es la media, \sum indica la suma, x_i son los valores individuales y n es el número total de valores.

Mediana La mediana es el valor medio de un conjunto de datos ordenados. Si el número de observaciones es impar, la mediana es el valor central. Si es par, es el promedio de los dos valores centrales.

Moda La moda es el valor o valores que aparecen con mayor frecuencia en un conjunto de datos.

Desviación Estándar

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (A.4)$$

Donde σ es la desviación estándar, x_i son los valores individuales y \bar{x} es la media aritmética.

A.2. Análisis Exploratorio de Datos

El Análisis Exploratorio de Datos (EDA, por sus siglas en inglés) es un enfoque estadístico utilizado para descubrir patrones, identificar anomalías, explorar estructuras de datos y probar hipótesis. Es una etapa preliminar en el análisis de datos que se realiza antes del modelado estadístico formal.

A.2.1. Objetivos del EDA

- Comprender la distribución de los datos.
- Maximizar el insight de un conjunto de datos.
- Descubrir estructuras subyacentes.
- Detectar valores atípicos y anomalías.
- Probar suposiciones.
- Desarrollar modelos parsimoniosos.
- Determinar relaciones óptimas entre variables.

A.2.2. Técnicas de EDA

Análisis univariado Se centra en una sola variable. Incluye técnicas como histogramas, boxplots y estadísticas descriptivas como la media y la desviación estándar.

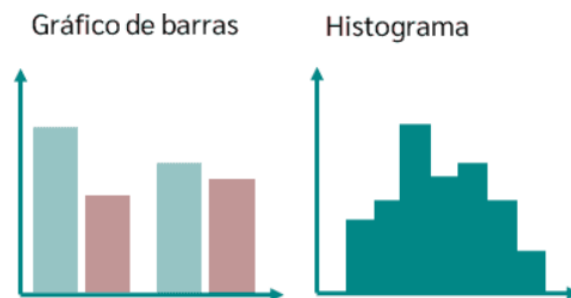
Análisis bivariado Explora la relación entre dos variables. Utiliza gráficos de dispersión, correlaciones y pruebas de hipótesis para evaluar la asociación.

Análisis multivariado Involucra el análisis de más de dos variables al mismo tiempo. Se utilizan técnicas como el análisis de componentes principales PCA y el análisis factorial.

A.2.3. Herramientas de Visualización

Gráficos de barras y histogramas Para variables categóricas y cuantitativas, respectivamente como se muestra en la figura A.1.

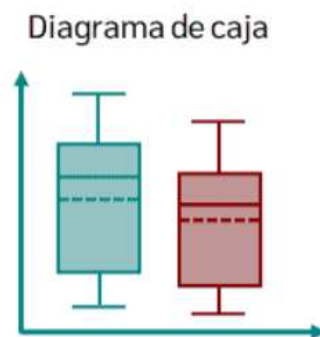
Figura A.1: Comparación Gráficos de barras y histogramas



Fuente: Adaptado de «Tutorial sobre gráficos» por DataTab (2024).

Diagrama de caja (Boxplots) Útiles para visualizar la distribución y detectar valores atípicos mostrado en la figura A.2.

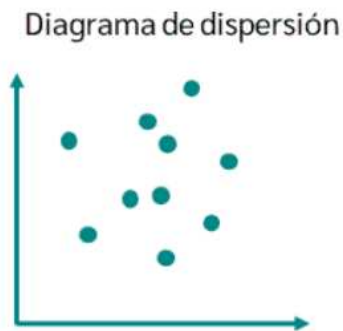
Figura A.2: Diagramas de Cajas



Fuente: Adaptado de «Tutorial sobre gráficos» por DataTab (2024).

Gráficos de dispersión (Scatter plots) Para examinar la relación entre variables continuas, en la figura A.3.

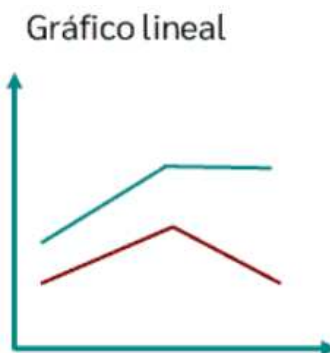
Figura A.3: Gráficos de dispersión



Fuente: Adaptado de «Tutorial sobre gráficos» por DataTab (2024).

Gráficos de líneas Para visualizar tendencias y patrones a lo largo del tiempo (figura A.4).

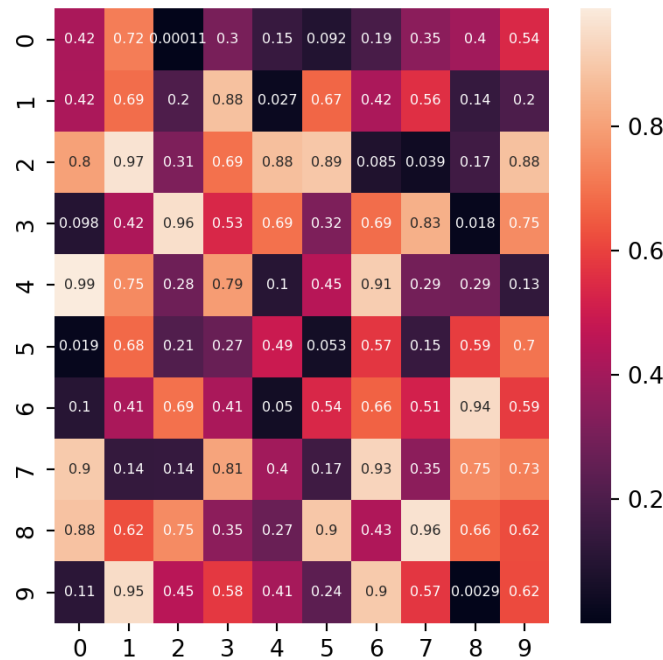
Figura A.4: Gráficos de líneas



Fuente: Adaptado de «Tutorial sobre gráficos» por DataTab (2024).

Mapas de Calor Utilizados para visualizar la matriz de correlación entre variables. Los colores intensos representan correlaciones fuertes, como se muestra en la figura A.5.

Figura A.5: Mapa de Calor



Fuente: Adaptado de «Tutorial sobre gráficos» por DataTab (2024).

A.2.4. Creación de Mapas de Calor y Gráficas de Dispersión

Para crear un mapa de calor o una gráfica de dispersión en Python se puede utilizar la biblioteca Seaborn, que es una interfaz de alto nivel para Matplotlib:

```
1 import seaborn as sns
2 import numpy as np
3 import matplotlib.pyplot as plt
4 # Generar datos aleatorios para el ejemplo
5 np.random.seed(0)
6 data = np.random.rand(10, 12)
7 # Crear el mapa de calor
8 sns.heatmap(data)
9 plt.show()
```

A.2.5. Estadísticas Descriptivas

- **Medidas de tendencia central:** Media (μ), mediana (Med) y moda (Mo).
- **Medidas de dispersión:** Rango, varianza (σ^2) y desviación estándar (σ).
- **Formas de distribución:** Asimetría y curtosis.

A.2.6. Detección de Valores Atípicos

Los valores atípicos son observaciones que se desvían significativamente del resto de los datos. Se pueden detectar utilizando:

- **Método de la desviación estándar:** Observaciones que caen fuera de 3σ de la media.
- **Método del rango intercuartílico (IQR):** Observaciones que están por debajo de $Q1 - 1,5 \times IQR$ o por encima de $Q3 + 1,5 \times IQR$.

A.2.7. Correlación y Causalidad

La correlación mide la relación lineal entre dos variables. Se representa con el coeficiente de correlación de Pearson (r), que varía entre -1 y 1. Sin embargo, correlación no implica causalidad.

El EDA es un paso crítico en el análisis de datos que proporciona una comprensión profunda de las características y relaciones en un conjunto de datos. Facilita la toma de decisiones informadas y la construcción de modelos analíticos sólidos.

A.3. Fundamentos de Machine Learning

El Machine Learning (ML) es una rama de la Inteligencia Artificial que se centra en el desarrollo de algoritmos que permiten a las máquinas aprender de los datos y mejorar su desempeño con el tiempo sin ser explícitamente programadas para cada tarea.

A.3.1. Tipos de Aprendizaje en Machine Learning

- **Aprendizaje Supervisado:** Los modelos predicen una salida basada en entradas etiquetadas. Se utilizan para clasificación y regresión.
- **Aprendizaje No Supervisado:** Los modelos identifican patrones y estructuras en datos no etiquetados. Se utilizan para clustering y reducción de dimensionalidad.
- **Aprendizaje por Refuerzo:** Los modelos aprenden a tomar decisiones a través de recompensas y penalizaciones.

A.3.2. Modelos y Algoritmos Comunes

- **Regresión Lineal:**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (\text{A.5})$$

Donde y es la variable dependiente, β_0 es el término de intercepción, β_i son los coeficientes de las variables independientes x_i , y ϵ es el error aleatorio.

- **Clasificación:** Algoritmos como Árboles de Decisión, K-Vecinos más Cercanos (KNN) y Máquinas de Vectores de Soporte (SVM).
- **Redes Neuronales y Deep Learning:** Consisten en capas de neuronas artificiales

que procesan la información de manera similar al cerebro humano.

$$a^{(l+1)} = \sigma(W^{(l)}a^{(l)} + b^{(l)}) \quad (\text{A.6})$$

Donde $a^{(l+1)}$ es la activación de la capa siguiente, $W^{(l)}$ y $b^{(l)}$ son los pesos y sesgos de la capa actual, y σ es la función de activación.

- **Clustering:** Algoritmos como K-Means y Clustering Jerárquico que agrupan datos basados en similitudes.
- **Reducción de Dimensionalidad:** Técnicas como Análisis de Componentes Principales (PCA) y t-SNE que reducen la cantidad de variables considerando solo las más significativas.

A.3.3. Evaluación de Modelos

Para evaluar la efectividad de un modelo de Machine Learning, se utilizan métricas como:

- **Para clasificación:** Precisión, Recall, F1-Score y ROC Curve.
- **Para regresión:** Error Cuadrático Medio (MSE), Error Absoluto Medio (MAE) y Coeficiente de Determinación (R^2).

A.3.4. Procesamiento de Lenguaje Natural

El Procesamiento de Lenguaje Natural (NLP, Natural Language Processing) es un subcampo del Machine Learning que se enfoca en la interacción entre computadoras y lenguaje humano. Incluye tareas como traducción automática, análisis de sentimientos y reconocimiento de voz.

El Machine Learning es una herramienta poderosa que está transformando la manera

en que interactuamos con la tecnología y analizamos grandes volúmenes de datos. Su aplicación abarca desde sistemas de recomendación hasta vehículos autónomos y diagnósticos médicos.

A.4. Aplicaciones Prácticas con Python

Python es un lenguaje de programación ampliamente utilizado en estadística y machine learning debido a su sintaxis clara y legible, así como a su rica colección de bibliotecas especializadas.

A.4.1. Bibliotecas de Python para Estadística y ML

- **NumPy:** Proporciona soporte para arrays y matrices, junto con una colección de funciones matemáticas para operar con estos objetos.
- **Pandas:** Ofrece estructuras de datos y herramientas para la manipulación y análisis efectivo de datos.
- **Matplotlib:** Una biblioteca de gráficos que permite la creación de figuras estáticas, animadas e interactivas en Python.
- **Seaborn:** Basada en Matplotlib, facilita la generación de gráficos estadísticos informativos y atractivos.
- **SciPy:** Utilizada para cálculos científicos y técnicos.
- **Scikit-learn:** Biblioteca para machine learning que ofrece herramientas simples y eficientes para análisis predictivos de datos.
- **TensorFlow y Keras:** Para redes neuronales y deep learning, permitiendo la construcción y entrenamiento de modelos avanzados.

A.4.2. Casos de Estudio y Ejemplos de Código

Análisis de Datos con Pandas

```
1 import pandas as pd
2 df = pd.read_csv('datos.csv')
3 print(df.head())
4 print(df.describe())
5
```

Visualización de Datos con Matplotlib

```
1 import matplotlib.pyplot as plt
2 df['columna'].hist(bins=50)
3 plt.show()
4
```

Modelo de Regresión con Scikit-learn

```
1 from sklearn.linear_model import LinearRegression
2 modelo = LinearRegression()
3 modelo.fit(X_train, y_train)
4 y_pred = modelo.predict(X_test)
5
```

Red Neuronal con Keras

```
1 from keras.models import Sequential
2 from keras.layers import Dense
3 modelo = Sequential()
4 modelo.add(Dense(12, input_dim=8, activation='relu'))
5 modelo.add(Dense(1, activation='sigmoid'))
6 modelo.compile(loss='binary_crossentropy', optimizer='adam',
7               metrics=['accuracy'])
8 modelo.fit(X, y, epochs=150, batch_size=10)
9
```

A.4.3. Mejores Prácticas y Consideraciones Éticas

Al trabajar con estadística y machine learning, es crucial seguir las mejores prácticas:

- Validar modelos con datos de prueba.
- Evitar el sobreajuste mediante técnicas de regularización.
- Considerar la privacidad de los datos y obtener consentimiento para su uso.
- Ser transparente sobre los métodos utilizados y los resultados obtenidos.

Python, junto con sus bibliotecas, proporciona un entorno robusto y flexible para llevar a cabo análisis estadísticos y desarrollar modelos de machine learning. Estas herramientas permiten a los investigadores y profesionales explorar datos, obtener insights y construir sistemas predictivos con eficacia y precisión.

APLICACIÓN DEL MÉTODO MACHINE LEARNING EN LA TOMA DE DECISIONES PARA OPTIMIZAR LA PERFORACIÓN DE POZOS

Este trabajo detalla el desarrollo de un modelo de *Machine Learning (ML)* para optimizar la Tasa de Penetración *Rate of Penetration (ROP)* en la perforación de pozos petroleros. El enfoque se centra en la implementación y comparación de algoritmos avanzados de ML, específicamente *Gradient Boosting Machine (GBM)*, *eXtreme Gradient Boosting (xGBoost)* y *Random Forest (RF)*, con el objetivo de mejorar la eficiencia operativa y reducir costos.

En particular, el algoritmo xGBoost mostró un rendimiento superior, seguido por Random Forest y GBM, indicando su potencial para implementarse en operaciones de perforación reales.

Estos hallazgos sugieren que la integración de modelos predictivos basados en ML puede resultar en una planificación más precisa y una ejecución más eficiente de la perforación de pozos petroleros. Las conclusiones del estudio enfatizan la importancia de la selección de algoritmos y la calidad de los datos en la precisión de las predicciones de la ROP.

Además, se recomienda realizar pruebas adicionales en un entorno de producción para validar la eficacia de los modelos en tiempo real y explorar su integración con sistemas de control automatizado. Este trabajo no solo contribuye al campo académico proporcionando una metodología robusta para la optimización de la ROP, sino que también ofrece una perspectiva aplicable para la mejora continua en las operaciones de perforación.

Luis Erwin Condori Mamani

luiserwinc@gmail.com

Cel.: 75858680



2024-TTES-462-D-1

**DIRECCIÓN DE DERECHO DE AUTOR
Y DERECHOS CONEXOS
RESOLUCIÓN ADMINISTRATIVA NRO. 1-1428/2024
La Paz, 10 de mayo de 2024**

VISTOS:

La solicitud de Inscripción de Derecho de Autor presentada en fecha **06 de mayo de 2024**, por **LUIS ERWIN CONDORI MAMANI** con **C.I. N° 9884935 LP**, con número de trámite **DA 789/2024**, señala la pretensión de inscripción del Trabajo Dirigido titulado: **"APLICACIÓN DEL MÉTODO MACHINE LEARNING EN LA TOMA DE DECISIONES PARA OPTIMIZAR LA PERFORACIÓN DE POZOS"**, cuyos datos y antecedentes se encuentran adjuntos y expresados en el Formulario de Declaración Jurada.

CONSIDERANDO:

Que, en observación al Artículo 4º del Decreto Supremo N° 27938 modificado parcialmente por el Decreto Supremo N° 28152 el *"Servicio Nacional de Propiedad Intelectual SENAPI, administra en forma desconcentrada e integral el régimen de la Propiedad Intelectual en todos sus componentes, mediante una estricta observancia de los regímenes legales de la Propiedad Intelectual, de la vigilancia de su cumplimiento y de una efectiva protección de los derechos de exclusiva referidos a la propiedad industrial, al derecho de autor y derechos conexos; constituyéndose en la oficina nacional competente respecto de los tratados internacionales y acuerdos regionales suscritos y adheridos por el país, así como de las normas y regímenes comunes que en materia de Propiedad Intelectual se han adoptado en el marco del proceso andino de integración"*.

Que, el Artículo 16º del Decreto Supremo N° 27938 establece *"Como núcleo técnico y operativo del SENAPI funcionan las Direcciones Técnicas que son las encargadas de la evaluación y procesamiento de las solicitudes de derechos de propiedad intelectual, de conformidad a los distintos regímenes legales aplicables a cada área de gestión"*. En ese marco, la Dirección de Derecho de Autor y Derechos Conexos otorga registros con carácter declarativo sobre las obras del ingenio cualquiera que sea el género o forma de expresión, sin importar el mérito literario o artístico a través de la inscripción y la difusión, en cumplimiento a la Decisión 351 Régimen Común sobre Derecho de Autor y Derechos Conexos de la Comunidad Andina, Ley de Derecho de Autor N° 1322, Decreto Reglamentario N° 23907 y demás normativa vigente sobre la materia.

Que, la solicitud presentada cumple con: el Artículo 6º de la Ley N° 1322 de Derecho de Autor, el Artículo 26º inciso a) del Decreto Supremo N° 23907 Reglamento de la Ley de Derecho de Autor, y con el Artículo 4º de la Decisión 351 Régimen Común sobre Derecho de Autor y Derechos Conexos de la Comunidad Andina.

Que, de conformidad al Artículo 18º de la Ley N° 1322 de Derecho de Autor en concordancia con el Artículo 18º de la Decisión 351 Régimen Común sobre Derecho de Autor y Derechos Conexos de la Comunidad Andina, referentes a la duración de los Derechos Patrimoniales, los mismos establecen que: *"la duración de la protección concedida por la presente ley será para toda la vida del autor y por 50 años después de su muerte, a favor de sus herederos, legatarios y cesionarios"*

Que, se deja establecido en conformidad al Artículo 4º de la Ley N° 1322 de Derecho de Autor, y Artículo 7º de la Decisión 351 Régimen Común sobre Derecho de Autor y Derechos Conexos de la Comunidad Andina que: *"...No son objeto de protección las ideas contenidas en las obras literarias, artísticas, o el contenido ideológico o técnico de las obras científicas ni su aprovechamiento industrial o comercial"*

Que, el artículo 4, inciso e) de la ley N° 2341 de Procedimiento Administrativo, instituye que: *"... en la relación de los particulares con la Administración Pública, se presume el principio de buena fe. La confianza, la cooperación y la lealtad en la actuación de los servidores públicos y de los"*



ciudadanos ...", por lo que se presume la buena fe de los administrados respecto a las solicitudes de registro y la declaración jurada respecto a la originalidad de la obra.

POR TANTO:

El Director de Derecho de Autor y Derechos Conexos sin ingresar en mayores consideraciones de orden legal, en ejercicio de las atribuciones conferidas.

RESUELVE:

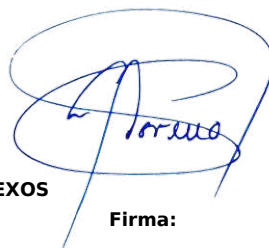
INSCRIBIR en el Registro de Tesis, Proyectos de Grado, Monografías y Otras Similares de la Dirección de Derecho de Autor y Derechos Conexos, el Trabajo Dirigido titulado: **"APLICACIÓN DEL MÉTODO MACHINE LEARNING EN LA TOMA DE DECISIONES PARA OPTIMIZAR LA PERFORACIÓN DE POZOS"** a favor del autor y titular: **LUIS ERWIN CONDORI MAMANI** con **C.I. N° 9884935 LP** bajo el seudónimo **LUISCM**, quedando amparado su derecho conforme a Ley, salvando el mejor derecho que terceras personas pudieren demostrar.

Regístrese, Comuníquese y Archívese.

CASA/Im

Firmado Digitalmente por:

Servicio Nacional de Propiedad Intelectual - SENAPI
CARLOS ALBERTO SORUCO ARROYO
DIRECTOR DE DERECHO DE AUTOR Y DERECHOS CONEXOS
LA PAZ - BOLIVIA



Firma:



f5toe9F17Df96j

PARA LA VALIDACIÓN DEL PRESENTE DOCUMENTO INGRESAR A LA PÁGINA WEB www.senapi.gob.bo/verificacion Y COLOCAR CÓDIGO DE VERIFICACIÓN O ESCANEAR CÓDIGO QR.



Oficina Central - La Paz
Av. Montes, N° 515,
entre Esq. Uruguay y
C. Batallón Illimani.
Telfs.: 2115700
2119276 - 2119251

Oficina - Santa Cruz
Av. Uruguay, Calle
prolongación Quijarro,
N° 29, Edif. Bicentenario.
Telfs.: 3121752 - 72042936

Oficina - Cochabamba
Calle Bolívar, N° 737,
entre 16 de Julio y Antezana.
Telfs.: 4141403 - 72042957

Oficina - El Alto
Av. Juan Pablo II, N° 2560
Edif. Multicentro El Ceibo
Ltda. Piso 2, Of. 5B,
Zona 16 de Julio.
Telfs.: 2141001 - 72043029

Oficina - Chuquisaca
Calle Kilómetro 7, N° 366
casi esq. Urriagoitia,
Zona Parque Bolívar.
Telf.: 72005873

Oficina - Tarija
Av. La Paz, entre
Calles Ciro Trigo y Avaroa
Edif. Santa Clara, N° 243.
Telf.: 72015286

Oficina - Oruro
Calle 6 de Octubre, N° 5837,
entre Ayacucho
y Junín, Galería Central,
Of. 14.
Telf.: 67201288

Oficina - Potosí
Av. Villazón entre calles
Wenceslao Alba y San Alberto,
Edif. AM. Salinas N° 242,
Primer Piso, Of. 17.
Telf.: 72018160

