ESCUELA DE NEGOCIOS

MAESTRÍA EN INTELIGENCIA DE NEGOCIOS Y CIENCIA DE DATOS

TÍTULO DE LA INVESTIGACIÓN

Modelo Predictivo de Producción de Hidrocarburos en el Campo Sacha operado por EP PETROECUADOR: Integración de Datos de Producción y Pruebas de Pozos

Profesor

Mario Salvador González

Autor

Gabriel Marcelo Pérez Padilla

2024

Resumen

La investigación "Modelo Predictivo de Producción de Hidrocarburos en el Campo Sacha operado por EP PETROECUADOR " se centra en elaborar un modelo predictivo para estimar la producción futura de crudo, agua y gas en el Campo Sacha. Utilizando datos de los sistemas TOW y LOWIS, y empleando técnicas avanzadas de análisis y machine learning, el estudio busca optimizar la eficiencia en la extracción de hidrocarburos. Se analizaron diversas metodologías predictivas, desde modelos lineales básicos hasta complejos modelos de ensamble como Random Forest. Los resultados demuestran que el modelo Random Forest ofrece una gran exactitud en las predicciones, facilitando una planificación y gestión más eficiente de los recursos. Este enfoque proporciona a EP PETROECUADOR una herramienta para mejorar la toma de decisiones, reducir riesgos y maximizar rentabilidad operativa del campo.

Abstract

The research "Predictive Model for Hydrocarbon Production in the Sacha Field Operated by EP PETROECUADOR" aims to develop a predictive model to estimate future production of oil, water, and gas in the Sacha Field. Using data from TOW and LOWIS systems and employing advanced analytical and machine learning techniques, the study seeks to optimize hydrocarbon extraction efficiency. Various predictive methodologies were analyzed, from basic linear models to complex ensemble models like Random Forest. Results show that the Random Forest model provides high prediction accuracy, enabling better resource planning and management. This approach offers EP PETROECUADOR a valuable tool for improving decision-making, reducing risks, and maximizing the operational profitability of the field.

Índice

Índice de Figuras

1. Introducción

La industria petrolera se enfrenta a una serie de retos constantes, entre los que se encuentran maximizar la eficiencia en la extracción de hidrocarburos, reducir costos operativos y minimizar riesgos asociados (World Petroleum Council, 2019). En este contexto, EP PETROECUADOR, la empresa pública encargada de exploración y extracción y comercialización de hidrocarburos en Ecuador, busca optimizar sus operaciones mediante el uso de tecnologías avanzadas y modelos predictivos que mejoren la toma de decisiones.

El Campo Sacha, ubicado en el Bloque 60 en la Provincia de Orellana y operado por EP PETROECUADOR, es uno de los yacimientos más importantes del país. La capacidad de prever con precisión la producción futura de crudo, agua y gas en este campo es esencial para la planificación estratégica y la administración eficiente de los recursos. Para abordar esta necesidad, esta investigación se centra en desarrollar un modelo predictivo que integre datos históricos de producción y resultados de pruebas de pozos.

La literatura existente en ingeniería de yacimientos petrolíferos ofrece una base sólida para comprender las diferentes metodologías utilizadas en la predicción de la producción de hidrocarburos, incluyendo modelos analíticos, numéricos y empíricos (Hoteit & Al-Kaabi, 2016). La integración de estos enfoques con técnicas modernas de análisis de datos y machine learning puede proporcionar estimaciones más precisas y útiles para la industria petrolera.

Este proyecto no solo busca mejorar la precisión de las predicciones de producción, sino también ofrecer una herramienta práctica para EP PETROECUADOR, permitiendo una gestión más efectiva del Campo Sacha y la opción de replicar en el resto de campos operados por la empresa. Al analizar y modelar datos de producción y pruebas de pozos, se pretende identificar tendencias, anomalías y relaciones clave entre variables que impacten la eficiencia operativa y la rentabilidad. Con una planificación más informada y una mejor comprensión del comportamiento del yacimiento, EP PETROECUADOR podrá optimizar sus estrategias de producción, maximizar la recuperación de hidrocarburos y minimizar los costos y riesgos asociados.

2. Revisión de la literatura relacionada al problema

2.1. Discusión de la literatura académica en relación al objeto de estudio

2.1.1. Modelado de yacimientos petrolíferos

La literatura sobre ingeniería de yacimientos petrolíferos proporciona una base sólida para comprender las diferentes metodologías utilizadas en la predicción de la producción de hidrocarburos. Se han desarrollado diversos enfoques, como modelos analíticos, numéricos y empíricos, para representar el comportamiento de los yacimientos y predecir la producción futura. (Caudle, 2020).

Se destacan los avances significativos en la comprensión y modelado de yacimientos petrolíferos. Se observa una evolución desde enfoques más simples, como los modelos analíticos, hacia métodos más sofisticados, que permiten una representación más precisa del comportamiento de los yacimientos. El progreso en tecnologías emergentes y la demanda de una predicción más precisa del comportamiento de los yacimientos han generado un cambio importante en la manera en que se enfoca la investigación y administración de los recursos petroleros. (Ertekin, 2005).

Es importante resaltar la relevancia de prever la producción de hidrocarburos para una planificación y administración óptima de los campos petrolíferos. Los modelos desarrollados permiten estimar la producción futura y proporcionan información esencial en el proceso de tomar decisiones dentro de la industria petrolera (Economides & Nolte, 2000). Esta información se utiliza para:

- Evaluar la viabilidad económica de un proyecto petrolífero.
- Elaborar el plan de producción y expansión del campo.
- Optimizar la operación del campo para maximizar la recuperación de hidrocarburos.
- Minimizar los costos y riesgos asociados a la producción.

2.2. Análisis de datos de producción

El estudio incluye datos históricos de crudo, agua y gas. La literatura sobre este tema ofrece una variedad de técnicas para explorar y entender los patrones en estos datos, incluyendo métodos estadísticos, análisis de series temporales, métodos de extracción de datos y automatización del aprendizaje (Clark & Frenkel, 2019). Estas técnicas permiten identificar tendencias, anomalías y relaciones entre variables, lo cual, a su vez, contribuye a tomar mejores decisiones en la gestión de los campos petrolíferos.

Se puede destacar la importancia del análisis de datos de producción en la industria petrolera para comprender el rendimiento de los campos petrolíferos e identificar tendencias y patrones sobre la gestión de los activos.

La literatura revisada muestra una amplia gama de métodos disponibles para examinar los datos de producción, que incluyen:

- Métodos estadísticos: análisis descriptivo e inferencial (Dake, 1978).
- Análisis de series temporales: para reconocer patrones tendencias y ciclos, a lo largo del tiempo (Chatfield, 2004).
- Métodos de extracción de datos: para identificar patrones no evidentes en volúmenes grandes de datos (Han, Kamber, & Pei, 2011).
- Aprendizaje automático: para desarrollar modelos que pueden predecir y clasificar datos (James, Witten, Hastie, & Tibshirani, 2013).

Se puede discutir cómo el análisis de datos de producción se utiliza para optimizar las operaciones petroleras, incluida la identificación de áreas de bajo rendimiento que requieren intervención, la detección de problemas de producción temprana y la optimización de la eficiencia de la producción. (Economides & Nolte, 2000).

Es importante discutir también los desafíos asociados con el análisis de datos de producción, la necesidad de contextualizar los resultados dentro del conocimiento geológico y operativo del yacimiento, y la interpretación adecuada de los resultados derivados de diversas metodologías analíticas.

2.2.1. Factores influyentes en la producción

En los yacimientos petrolíferos convencionales, la producción de hidrocarburos está influenciada por diversos factores, como la geología del yacimiento, la presión del yacimiento, la saturación de fluidos, y el método de recuperación aplicado (por ejemplo, bombeo primario, secundario o terciario). La literatura sobre estos temas proporciona información crucial para entender cómo estos factores afectan la producción y cómo se pueden incorporar en modelos predictivos para optimizar la producción de hidrocarburos (Gupta & Pande, 2018).

Es fundamental destacar la importancia de comprender cómo diversos factores geológicos, operativos y ambientales afectan la producción de hidrocarburos en los yacimientos petrolíferos. Esta comprensión es crucial para optimizar la producción, maximizar el rendimiento de los campos petrolíferos y minimizar el impacto ambiental de la industria petrolera (Al-Dhahri, 2023).

2.2.2. Modelos de pronóstico y predicción

Existen numerosos estudios que han desarrollado modelos de pronóstico y predicción para estimar producción futura de crudo, agua y gas en campos petrolíferos. Estos modelos pueden variar desde métodos básicos que se fundamentan en regresión, series temporales, hasta modelos más complejos que utilizan redes neuronales, algoritmos genéticos o técnicas de IA para incrementar la precisión de las predicciones (Moghaddam & Rasouli, 2020).

Es esencial destacar la amplia gama de enfoques utilizados en la construcción de modelos de pronóstico y predicción en la industria petrolera. Desde modelos simples basados en regresión hasta modelos más complejos que emplean técnicas avanzadas de inteligencia artificial como redes neuronales y aprendizaje automático.

Los diferentes tipos de modelos ofrecen flexibilidad para adaptarse a diferentes contextos y requisitos específicos de la producción petrolera. Por ejemplo, los modelos de regresión pueden ser útiles para pronósticos a corto plazo con datos

históricos limitados, mientras que las técnicas de inteligencia artificial pueden ser más adecuadas para pronósticos a largo plazo o cuando hay grandes volúmenes de datos disponibles.

La literatura revisada resalta cómo los avances en técnicas de modelado, especialmente en inteligencia artificial, han llevado a mejorar la precisión de los perfiles de producción (Moghaddam & Rasouli, 2020). Estos modelos más avanzados pueden captar relaciones no lineales y patrones complejos en datos, permitiendo una mejor estimación de la producción futura.

Es importante discutir la importancia de la validación y calibración de los modelos de pronóstico y predicción utilizando datos históricos y datos de producción reales. Esto ayuda a mejorar la fiabilidad de las predicciones y proporciona confianza en los resultados obtenidos.

2.2.3. Optimización de la producción

Existe información valiosa sobre cómo maximizar la producción de hidrocarburos mientras se minimizan los costos y se cumplen los objetivos operativos y ambientales. Estos estudios pueden ofrecer perspectivas sobre estrategias de gestión de la producción que pueden complementar los modelos de predicción desarrollados. (Kiani & Pourafshary, 2019).

La aplicación de la inteligencia artificial (IA) va creciendo para mejorar la eficiencia en la extracción de petróleo, agua y gas. La IA puede ayudar a:

- Analizar grandes conjuntos de datos: La IA puede examinar grandes conjuntos de datos de producción, geológicos y operativos para detectar patrones y tendencias que no son fácilmente detectables por métodos tradicionales.
- Desarrollar modelos predictivos más precisos: La IA se puede utilizar para desarrollar modelos predictivos más precisos en la producción de petróleo, agua y gas.

- Optimizar las estrategias de producción: La IA se puede utilizar para optimizar las estrategias de producción, como la tasa de producción, la selección del método de recuperación y la planificación de pozos.

La IA en la industria petrolera aún se encuentra en sus primeras etapas, pero podría aumentar notablemente la eficiencia en la producción y la recuperación de hidrocarburos. **(**Christie, M. A., & Blunt, M. J., 2014).

2.3. Detalle de fuentes primarias y secundarias

2.3.1. Fuentes primarias

Las fuentes primarias utilizadas en este trabajo provienen de los sistemas TOW (Test of Wells) y LOWIS (Live Oil Well Information System) de EP PETROECUADOR. Estos sistemas proporcionan datos directos y originales sobre producción de crudo, agua y gas en el campo Sacha, ubicado en el Bloque 60 en la Provincia de Orellana. Los datos recopilados son específicos del año 2023 y son extraídos de pozos operados por EP PETROECUADOR. Estos datos incluyen resultados de pruebas de pozos, mediciones de producción y otros parámetros relevantes para la investigación sobre la predicción de producción en el campo Sacha.

2.3.2. Fuentes Secundarias

Además de las fuentes primarias mencionadas, este trabajo se apoya en una variedad de fuentes secundarias que proporcionan análisis, interpretaciones y contextos adicionales sobre el tema en estudio. Estas fuentes secundarias incluyen libros, artículos de revistas científicas, informes técnicos y documentos gubernamentales relacionados con la industria petrolera y la predicción de la producción de hidrocarburos. Entre estas fuentes se encuentran estudios académicos sobre ingeniería de yacimientos petrolíferos, análisis de datos de producción, modelado predictivo en la industria petrolera, así como informes de organizaciones gubernamentales y agencias reguladoras pertinentes a la industria petrolera en Ecuador.

3. Identificación del objeto de estudio

3.1. Problema de Investigación

La industria petrolera, incluyendo EP PETROECUADOR, enfrenta el desafío de maximizar la eficiencia en la extracción de hidrocarburos mientras se minimizan los costos y se garantiza la integridad de los pozos (Economides & Nolte, 2000). Lograr este equilibrio es crucial para la sostenibilidad económica y ambiental de la industria.

Es fundamental comprender la relación existente entre la producción de crudo, agua y gas, y los resultados a las pruebas realizadas en los pozos (Dake, 1978). Esta información permite tomar decisiones fundamentadas acerca de la operación de pozos, selección de la mejor técnica de producción y la frecuencia de las pruebas.

Existen diversas técnicas para mejorar la eficiencia en la extracción de hidrocarburos (Society of Petroleum Engineers, 2023).

3.2. Justificación del proyecto y su importancia para resolver el problema organizacional identificado

Este proyecto se alinea con los objetivos estratégicos de EP PETROECUADOR de incrementar la eficiencia en la producción de hidrocarburos, optimizar el uso de recursos y minimizar riesgos. (EP PETROECUADOR Plan Estratégico 2023-2027)

Este proyecto tiene como objetivo crear un modelo predictivo que pueda estimar la producción futura de crudo, agua y gas en el Campo Sacha con un alto grado de precisión. Esta información será importante en la toma de decisiones de EP PETROECUADOR, permitiéndole:

3.2.1. Mejora en la extracción de hidrocarburos

Al elaborar el modelo predictivo para estimar la producción futura de crudo, agua y gas en el Campo Sacha, operado por EP PETROECUADOR, se proporcionará

a la empresa una herramienta valiosa para optimizar la estrategia de producción y minimizar costos operativos.

3.2.2. Toma de decisiones informadas

Identificar la relación entre la producción de crudo, agua y gas y los resultados en las pruebas realizadas a los pozos es fundamental para la toma decisiones respecto a la operación de los pozos. Al analizar esta relación y desarrollar modelos predictivos precisos, se proporcionará a la Gerencia de Exploración y Producción de EP PETROECUADOR la información necesaria para seleccionar la mejor técnica de producción, planificar inversiones y gastos de manera eficiente, y minimizar riesgos para maximizar la rentabilidad del proyecto.

3.2.3. Mejora de la planificación y la gestión de activos

Adicionalmente, un modelo predictivo robusto y adaptable, permitirá al Gerente de Exploración y Producción y, por lo tanto, al Gerente del Activo Sacha planificar de manera efectiva las operaciones de extracción de hidrocarburos en el Campo. La capacidad de capturar tendencias y patrones en los datos históricos, incorporar información adicional de manera efectiva y actualizar el modelo con nuevos datos garantizará una planificación más precisa y una administración efectiva de los recursos de petróleo.

3.2.4. Reducción de riesgos y maximización de la rentabilidad

Al proporcionar predicciones precisas sobre la producción futura de crudo, agua y gas, permitirá a EP PETROECUADOR minimizar riesgos y maximizar la rentabilidad del proyecto en el Campo Sacha. Con una planificación más precisa, la empresa podrá optimizar su estrategia de producción y maximizar los beneficios económicos y ambientales de la extracción de hidrocarburos.

4. Planteamiento del problema

4.1. Problemática organizacional a ser estudiada

La problemática organizacional a ser estudiada radica en la necesidad de EP PETROECUADOR de maximizar la eficiencia en la extracción de hidrocarburos en el Campo Sacha. Esta necesidad surge en un contexto donde la optimización de la producción petrolera es crucial para la sostenibilidad económica y ambiental de la industria petrolera.

El problema es crítico para la organización porque afecta directamente su capacidad para operar de manera rentable y eficiente en el Campo Sacha, con la posibilidad de replicar al resto de Campos operados por EP PETROECUADOR.

La dificultad en la comprensión clara de la relación entre la producción y los resultados de las pruebas realizadas puede llevar a decisiones operativas subóptimas y a un uso ineficiente de los recursos.

La justificación para adoptar un enfoque analítico radica en la necesidad de EP PETROECUADOR de basar sus decisiones en datos sólidos y análisis rigurosos. Un enfoque analítico permitirá a la organización mejorar la comprensión de los modelos y tendencias en los registros de producción y pruebas de pozos, lo que a su vez ayudará a optimizar la estrategia de producción.

4.2. Objetivos generales del proyecto

- Elaborar un modelo predictivo que permita estimar la producción futura de crudo, agua y gas en el Campo Sacha, operado por EP PETROECUADOR.
- Capturar tendencias en los datos históricos de producción de crudo, agua y gas, así como en los resultados de las pruebas de los pozos, para mejorar la comprensión del comportamiento del yacimiento.
- Incorporar de manera efectiva la información adicional, como las características geológicas y operativas del yacimiento, en el modelo predictivo para mejorar su precisión y capacidad de adaptación.

4.3. Objetivos específicos del proyecto

- Analizar la relación entre los datos de producción de crudo, agua y gas y los resultados de las pruebas de los pozos.
- Reconocer patrones y tendencias en los datos que puedan sugerir posibles correlaciones entre la producción y las características de los pozos.
- Desarrollar un modelo predictivo para estimar la producción de crudo, agua y gas, utilizando técnicas analíticas avanzadas como análisis estadístico y aprendizaje automático.
- Evaluar la eficacia del modelo propuesto en la predicción de la producción y su utilidad en la optimización de las operaciones petroleras, comparando las proyecciones del modelo con los resultados reales de producción para analizar su robustez y precisión.

El modelo debe ser capaz de:

- Identificar tendencias presentes en los datos históricos.
- Incorporar la información adicional de manera efectiva.
- Ser robusto y adaptable a cambios en las condiciones del yacimiento.
- Ser fácil de usar y actualizar con nuevos datos.

5. Justificación y aplicación de la Metodología a utilizar

5.1. Recolección de datos

En esta fase se identificaron las bases de datos que comprenden, en primer lugar, los registros históricos de producción, además de los resultados de las pruebas ejecutadas en los pozos. Extraídos de los sistemas TOW (Test of Wells) y LOWIS (Live Oil Well Information System) de EP PETROECUADOR, estos datos proporcionan información directa y original sobre la producción en el campo Sacha durante el año 2023.

Se realizó la recopilación de datos utilizando un entorno de desarrollo en Jupyter, donde se cargaron dos bases de datos relevantes para el análisis cuantitativo: 'produccion_df' con 2.028.637 datos y 'test_df' con 605.151 datos. Estos archivos contienen datos de producción de crudo, agua y gas, y los resultados específicos de las pruebas realizadas, siendo seleccionados por su directa relevancia para los objetivos del proyecto.

Es importante destacar que la elección de estas bases de datos se alinea estrechamente con los objetivos del proyecto, garantizando la disponibilidad de información confiable y actualizada para la investigación sobre la predicción de producción en el campo Sacha.

5.2. Limpieza, preprocesamiento y/o transformación de datos.

Para garantizar la calidad de los datos, se realizaron las siguientes tareas específicas:

- Eliminación de registros duplicados en las bases de datos de producción y pruebas para evitar redundancias que podrían sesgar el análisis.
- Descarte de columnas irrelevantes que no contribuían aportando información significativa al modelo o que podrían introducir ruido en el análisis. Específicamente, se eliminaron las columnas 'OPERACION', 'ACTIVO' y 'BLOQUE_N' de la base de producción.
- Identificación de las columnas críticas en las dos bases de datos y eliminación de las filas que contenían valores nulos en estas columnas.

Para otras columnas, se aplicaron métodos estadísticos como la media o mediana agrupada por categorías específicas o interpolación lineal, con el fin de imputar los valores faltantes.

- Las columnas de fecha en las bases de datos de producción y pruebas se convirtieron al formato correcto utilizando pd.to_datetime. Esto es esencial para garantizar que las operaciones subsecuentes que involucren fechas se realicen correctamente.

5.3. Identificación y Descripción de Variables

Las variables se han dividido en predictoras y objetivo. Las variables predictoras son aquellas que el modelo utilizará para realizar sus predicciones, mientras que las variables objetivo son los valores que se intentarán predecir.

- En este caso, se han seleccionado como variables predictoras:

- **AMPS_A:** Amperaje del motor de la bomba (amperios)
- **TUBING_TMP:** Temperatura en la tubería (grados Fahrenheit)
- **CASING_PRESS:** Presión en la camisa del pozo (psi)
- **PBHP:** Presión del fondo del pozo (psi)
- **SALINIDAD:** Salinidad del fluido producido (partes por millón)
- **GAS_SP_GRAVITY:** Gravedad específica del gas
- **STAGE_COUNT:** Número de etapas de la bomba
- **POWER_KW:** Potencia del motor de la bomba (kilovatios)
- **TUBING_PRESS:** Presión en la tubería de producción (psi)
- **PUMP_TMP:** Temperatura de la bomba (grados Fahrenheit)
- **VOLTAGE:** Voltaje aplicado a la bomba
- **INTAKE_DEPTH:** Profundidad de la toma de la bomba (pies)
- **POZO:** Nombre del Pozo (categórica)
- **COMP_S_NAME:** Nombre de la completación del pozo (Categórica)
- **PUMP_TYPE:** Tipo de bomba utilizada (Categórica)
- **YEAR:** Año

- **MONTH:** Mes
- **DAY:** Día
- **DAYOFWEEK:** Día de la semana

- Se han definido como variables como objetivo:

- **PC_BPPD:** Producción bruta diaria de petróleo (barriles por día)
- **WAT:** Producción diaria de agua (barriles por día)
- **GAS**: Producción diaria de gas (barriles equivalentes)

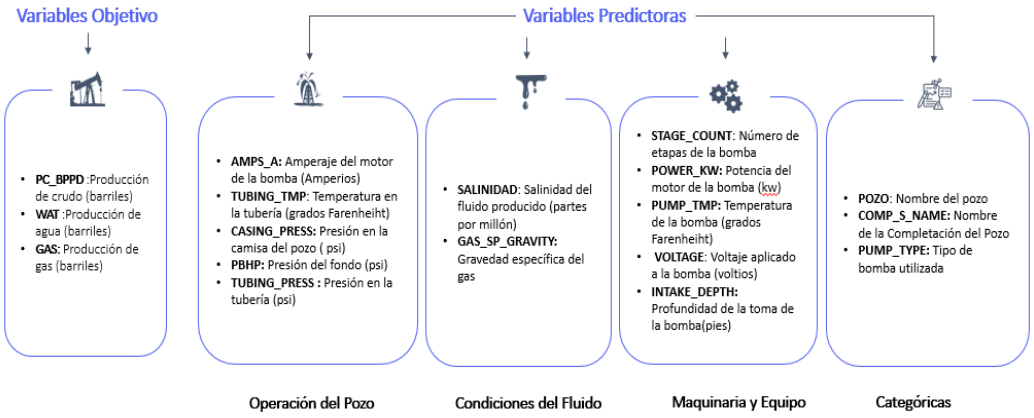

Figura 1. Clasificación de Variables

- Se realizó una fusión de los DataFrames de producción y pruebas utilizando como claves el nombre de la completación de los pozos ('COMP_S_NAME') y las fechas correspondientes. Esto permitió consolidar la información relevante en un solo DataFrame para un análisis más integrado.
- Se extrajeron características adicionales de la columna de fecha, como el año, mes, y día. Estas características podrían revelar patrones estacionales o tendencias a lo largo del tiempo que son valiosos para el análisis predictivo.
- Se aplicó codificación de etiquetas (Label Encoding) a las variables categóricas seleccionadas para convertirlas en un formato numérico que

los modelos de machine learning pueden interpretar. Se almacenaron los mapeos de LabelEncoder en un diccionario para su uso futuro, (Smith et al., 2018).

- Se eliminaron las columnas de fecha originales para evitar redundancias y se verificó que todas las variables predictoras estuvieran presentes en el DataFrame combinado.
- Se definieron las variables predictoras y objetivo, y se segmentó el conjunto de datos en entrenamiento y prueba. Se considera importante en el entrenamiento del modelo para evaluar su rendimiento de manera objetiva (Géron, 2017).

5.4. Visualización de Variables

Se han empleado diversas técnicas de visualización avanzadas para explorar, interpretar y evaluar los datos y el rendimiento del modelo predictivo.

5.4.1. Matriz de Correlación

Esta matriz es instrumental para discernir las relaciones lineales entre las variables. Los coeficientes de correlación, representados mediante un mapa de calor, proporcionan una visión clara de las posibles interdependencias, lo que es esencial para evitar la multicolinealidad en los modelos predictivos (VanderPlas, 2016).

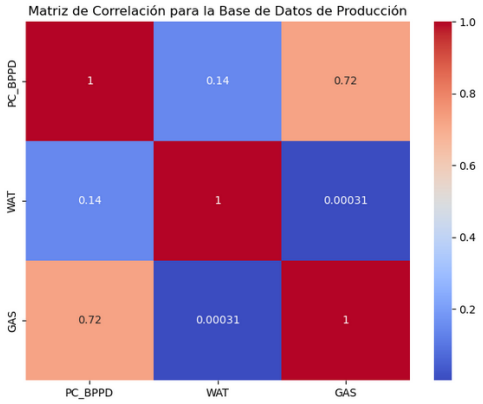

Figura 2. Matriz de Correlación Producción (crudo, agua y gas)

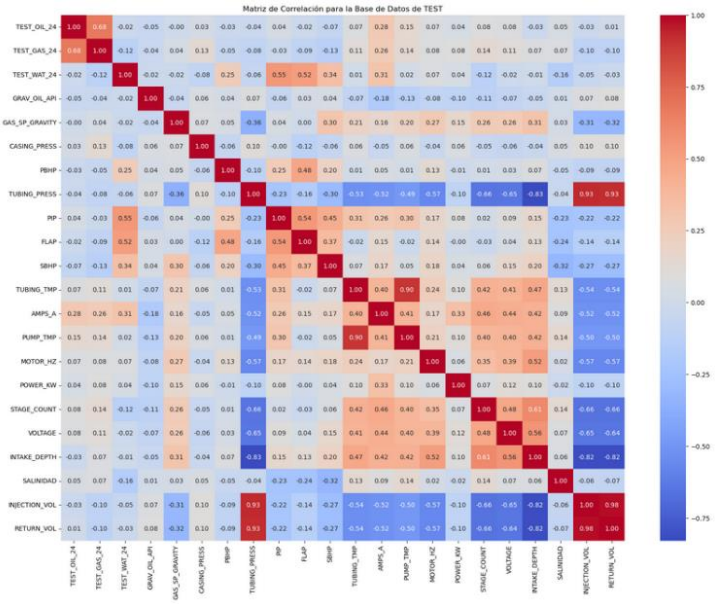

Figura 3. Matriz de Correlación Base Test

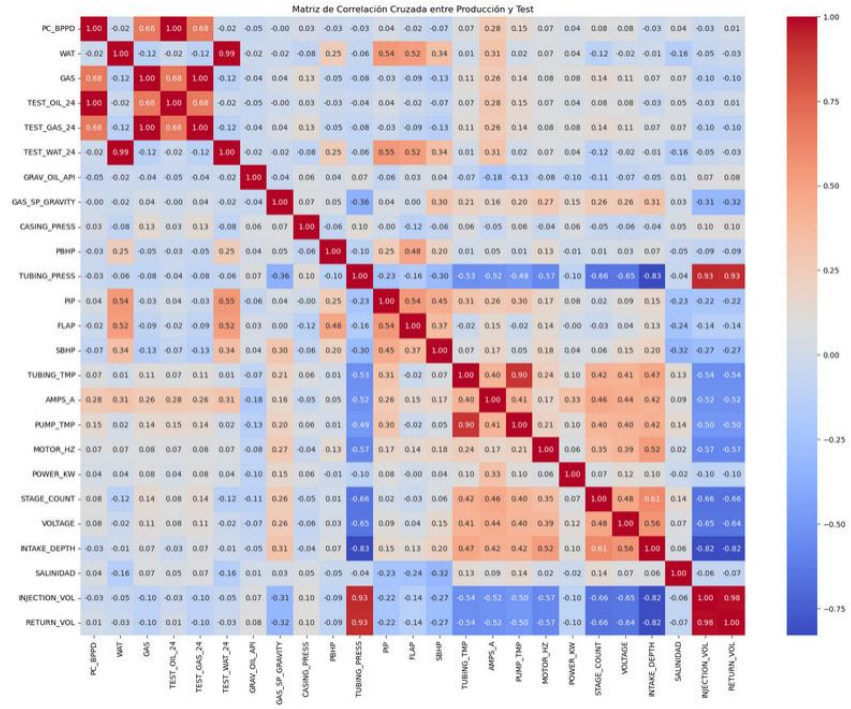

Figura 4. Matriz de Correlación Cruzada entre Base Producción y Test

5.4.2. Gráficos de Líneas Temporales

Se crearon gráficos de líneas para mostrar la producción promedio de crudo, agua y gas a lo largo del tiempo, agrupados por mes. Permiten detectar tendencias y patrones estacionales, así como para identificar posibles anomalías o eventos atípicos en la producción.

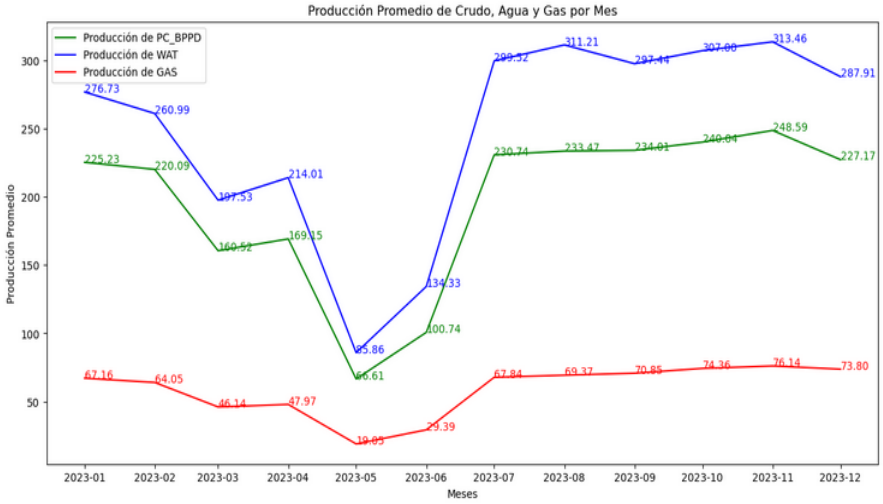

Figura 5. Producción Promedio Mensual

5.4.3. Boxplots

Se emplearon boxplots para comparar la distribución de la producción durante las pruebas.

Figura 6. Producción durante pruebas

EL diagrama de cajas y bigotes presenta cómo se distribuyen los datos e incluye detalles sobre la mediana, los cuartiles y posibles valores atípicos. La caja cubre el rango intercuartil (IQR), que comprende el 50% de los datos. El borde inferior de la caja indica el primer cuartil (Q1) y el borde superior corresponde al tercer cuartil (Q3). (McGill, Tukey, & Larsen, 1978).

La línea de la mediana en la caja del crudo es más alta que en las cajas del gas y del agua, esto indica que la mediana de la producción de crudo es mayor.

La caja del agua es más alta que la del crudo y el gas, esto indica una mayor variabilidad en la producción de agua. Los bigotes más largos también indicarían mayor variabilidad en el agua. Los puntos fuera de los bigotes en la caja del agua sugieren que hubo algunas pruebas con una producción de agua inusualmente alta o baja.

5.4.4. Comparación de Distribuciones

Se desarrolló una función personalizada para analizar la efectividad del modelo. Esta función genera gráficos de dispersión y series temporales, así como zooms en segmentos específicos de datos, permitiendo el análisis de la precisión del modelo. En la figura 8, se observarla distribución de los valores reales con los predichos de la variable de crudo, mostrando un buen rendimiento del modelo.

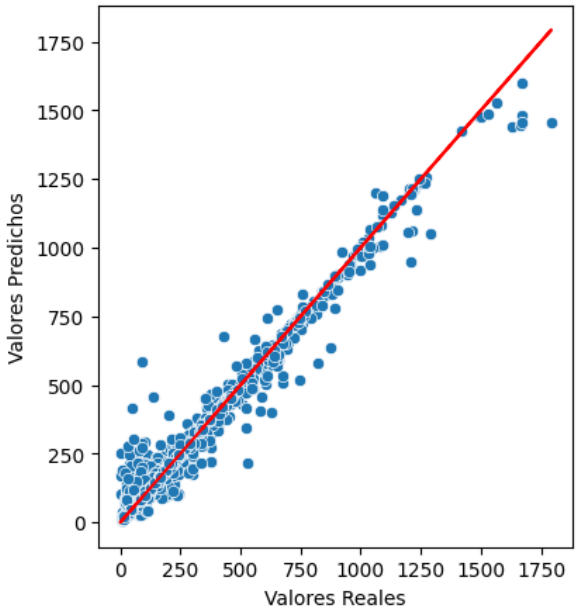

Figura 7. Scatter plot de valores reales vs predichos en crudo

5.4.5. Pairplot

Los pairplots muestran la relación entre cada par de variables en un DataFrame. En el caso de las figuras a continuación, se observan las relaciones existentes entre las variables predictoras seleccionadas (AMPS_A', 'TUBING_TMP', 'CASING_PRESS', 'PBHP', 'SALINIDAD', 'GAS_SP_GRAVITY) y las variables objetivo PC_BPPD (Crudo), WAT (Agua) y GAS.

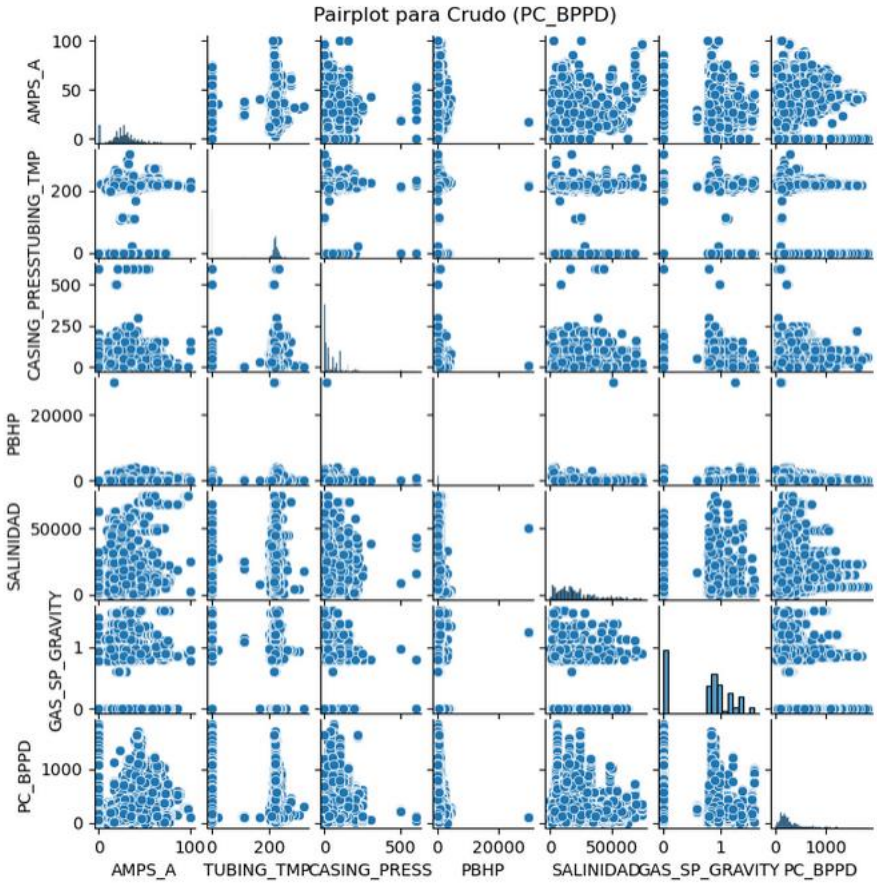

Figura 8. Pairplot para Crudo (PC_BPPD)

- Las distribuciones univariadas muestran la forma general de los datos para cada variable. Algunas variables como SALINIDAD y GAS_SP_GRAVITY parecen tener distribuciones más concentradas, mientras que otras como PC_BPPD están más dispersas.
- En general, no se observan patrones claros de correlación fuerte entre las variables predictoras y PC_BPPD. La mayoría de las relaciones son bastante dispersas, lo que sugiere que puede no haber una relación lineal fuerte entre estas variables y la producción de crudo.
- La falta de una correlación clara puede indicar que otros factores no incluidos en este análisis pueden estar influyendo en PC_BPPD o que las relaciones son más complejas y no lineales.

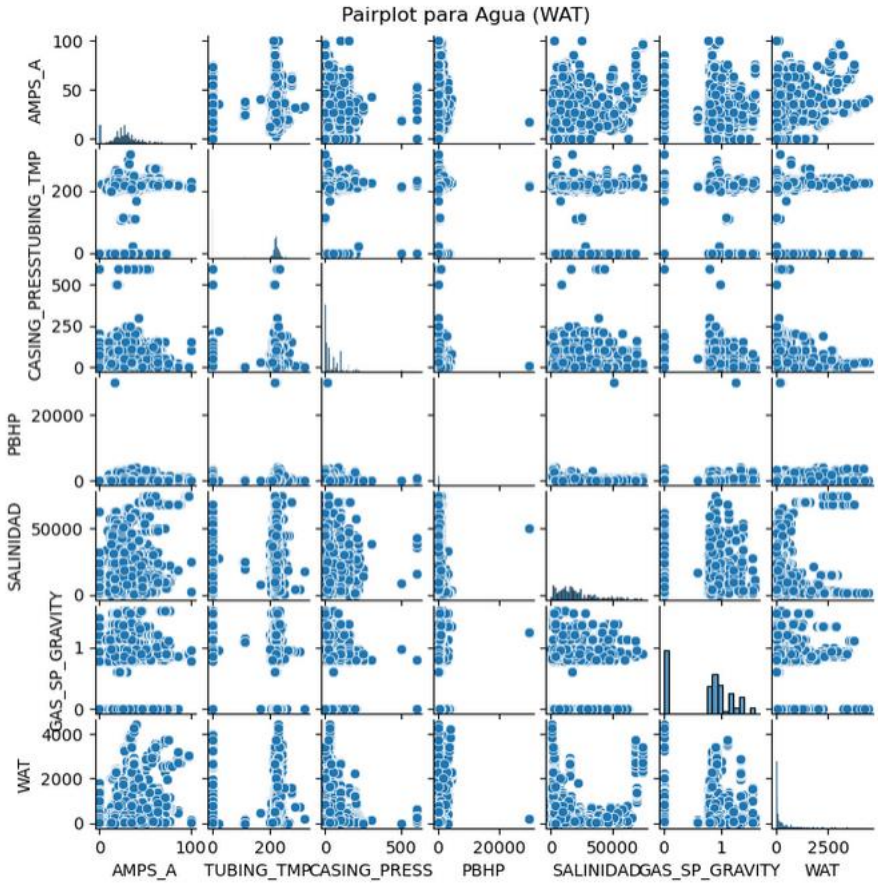

Figura 9. Pairplot para Agua (WAT)

- En general, no se observan patrones claros de correlación fuerte entre las variables predictoras y WAT. La mayoría de las relaciones son bastante dispersas, lo que sugiere que puede no haber una relación lineal fuerte entre estas variables y la cantidad de agua.
- La relación entre SALINIDAD y WAT muestra varios puntos concentrados en áreas específicas, lo que podría indicar una posible correlación en ciertas condiciones, pero en general, la relación parece ser débil.

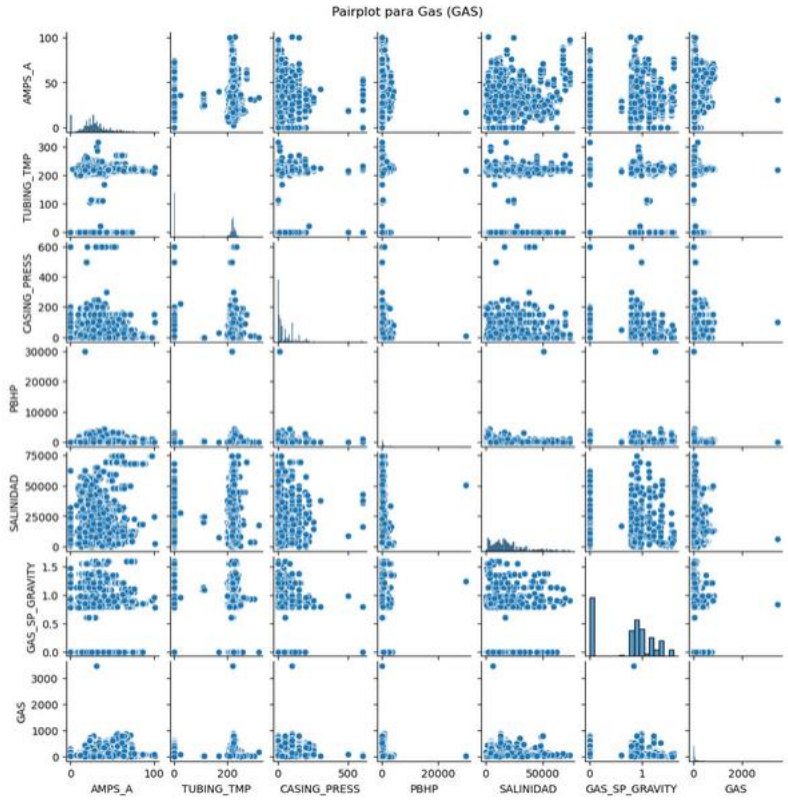

Figura 10. Pairplot para Gas

- La relación entre AMPS_A (amperios) y GAS muestra una nube de puntos dispersa, indicando una correlación débil o nula.
- La relación entre TUBING_TMP (temperatura del tubo) y GAS también muestra una dispersión considerable sin un patrón claro, sugiriendo una posible falta de correlación significativa.
- En general, no se observan patrones claros de correlación fuerte entre las variables predictoras y GAS. La mayoría de las relaciones son bastante dispersas, lo que sugiere que puede no haber una relación lineal fuerte entre estas variables y la cantidad de gas.

Con estos resultados, se decidió explorar modelos no lineales y técnicas de machine learning para desbloquear el potencial completo de los datos y así, capturar relaciones más complejas entre las variables.

5.5. Selección de Modelo Estadístico

Se ha seguido un enfoque metódico para identificar el modelo más eficaz, comenzando con técnicas de regresión lineal y progresando hacia modelos de ensamble más complejos.

La exploración comenzó con modelos lineales básicos, incluyendo Regresión Lineal y Regresión Ridge. A pesar de su simplicidad y rapidez de implementación, estos modelos no lograron capturar la complejidad de los datos, como se evidenció en sus altos valores de MSE y bajos coeficientes R^2. La regresión lineal, en particular, mostró un rendimiento insuficiente, lo que indicó la necesidad de modelos más sofisticados capaces de manejar relaciones no lineales y de alta dimensión.

Por consiguiente, La Regresión Lineal no se seleccionó para este proyecto debido a su capacidad limitada para gestionar la complejidad de los datos y su rendimiento inferior en comparación con modelos más avanzados, como el Random Forest.

Este último pudo ofrecer una mejor precisión predictiva y generalización al capturar de manera más efectiva las relaciones complejas entre las variables predictoras y las variables objetivo.

Con la intención de mejorar el rendimiento, se cambió hacia modelos basados en árboles, comenzando con un Árbol de Decisión. A través de la evaluación de residuos y la optimización de hiperparámetros mediante GridSearchCV, mejoró significativamente el MSE y el R^2. Sin embargo, la naturaleza propensa al sobreajuste de los árboles de decisión individuales llevó a considerar modelos de ensamble (Hastie, Tibshirani, & Friedman, 2009).

Los modelos de ensamble, como Random Forest y XGBoost, ofrecieron una mejora notable en la precisión predictiva. Estos modelos combinan las predicciones de múltiples árboles para contribuir a la generalización y minimizar la varianza. El Random Forest, en particular, demostró ser superior, alcanzando un MSE y un R^2 que superaban a los de XGBoost y los modelos lineales.

Se realizó una validación cruzada de varios modelos de machine learning. En esta evaluación, se utilizó el coeficiente de determinación R^2 como métrica principal para la evaluación del desempeño de los modelos.

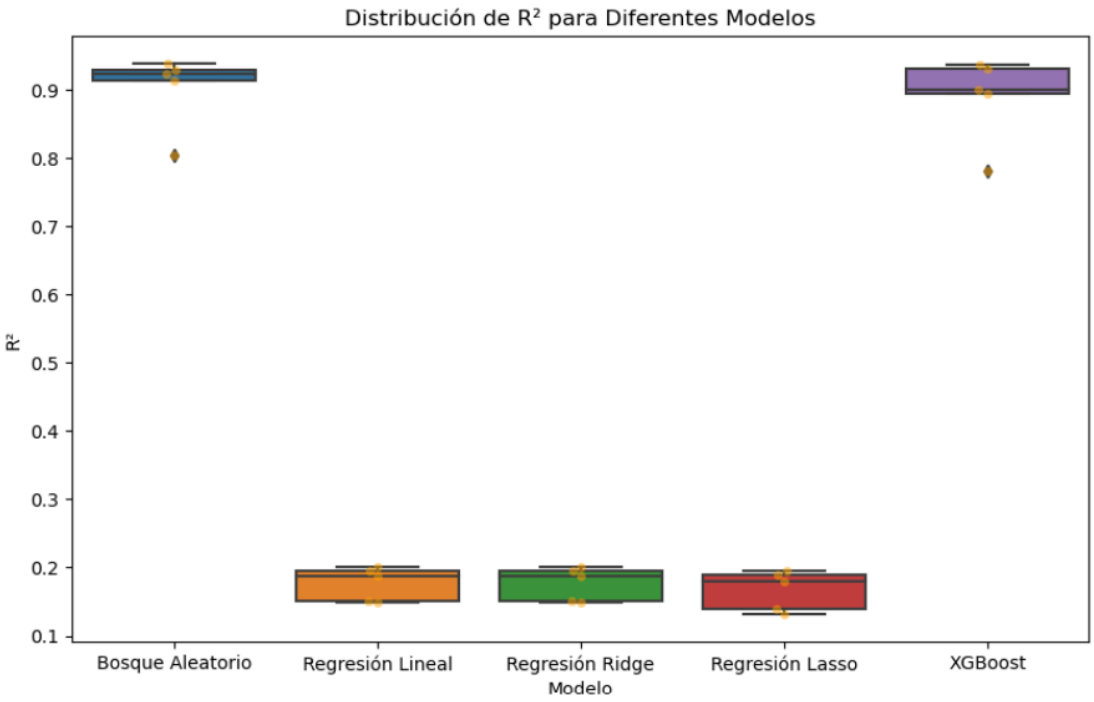

Figura 11. Validación cruzada diferentes Modelos

La figura 9 muestra la distribución de los puntajes R^2 para cada modelo. La línea central en cada caja representa la mediana de los puntajes R^2. Además, los extremos inferiores y superiores de las cajas representan el primer y tercer cuartil, respectivamente, lo que nos ofrece información sobre la dispersión de los puntajes R^2 alrededor de la mediana. Esta representación visual nos permite comparar fácilmente el rendimiento relativo de los diferentes modelos y entender la variabilidad en sus puntajes R^2 (Marsgr6, n.d.).

Con el Random Forest como modelo elegido, se procedió a una afinación más profunda de los hiperparámetros. La búsqueda exhaustiva reveló la configuración óptima que maximizaba la precisión predictiva, resultando en el mejor MSE y R^2 observados hasta el momento. Este modelo final no solo capturó la complejidad de los datos, sino que también mantuvo la capacidad de

generalizar bien a nuevos datos, como se confirmó mediante la validación cruzada.

5.5.1. Justificación de la Selección del Coeficiente de Determinación R^2 para la Evaluación de Modelos

El coeficiente de determinación R^2 se utiliza ampliamente como una métrica principal para la evaluación del desempeño de los modelos de regresión por varias razones clave, que justifican su uso en este contexto específico:

- **Medida de Bondad del Ajuste**

  El R^2 ofrece una medida clara de la calidad del ajuste de un modelo de regresión, al mostrar qué proporción de la variabilidad total en la variable de respuesta puede ser explicada por el modelo. En otras palabras, un R^2 elevado indica que el modelo refleja adecuadamente las tendencias y patrones subyacentes en los datos:

  - El R^2 cercano a 1 señala que el modelo explica casi toda la variabilidad en los datos.
  - EL R^2 cercano a 0 Indica que el modelo no captura la variabilidad de los datos mejor que un modelo que simplemente predice la media (Hastie, Tibshirani, & Friedman, 2009).

- **Comparación Relativa Entre Modelos**

  El R^2 facilita la comparación entre diferentes modelos de manera estandarizada. Al evaluar y comparar varios modelos (Regresión Lineal, Ridge, Árbol de Decisión, Random Forest, XGBoost), el R^2 ofrece un criterio común para determinar cuál modelo tiene un mejor desempeño en términos de ajuste a los datos:

  La figura 9, que muestra la distribución de los puntajes R^2 para cada modelo, permite comparar visualmente el rendimiento relativo de los modelos y la variabilidad en sus puntajes. Esta representación gráfica

ayuda a identificar el modelo que no solo tiene el mejor rendimiento promedio sino también la menor variabilidad, lo que es crucial para la generalización a nuevos datos.

- **Balance entre Complejidad y Precisión**

  El uso del R^2 junto con otras métricas como el MSE proporciona un balance entre la precisión del modelo y su complejidad. Mientras que el MSE mide el error absoluto del modelo, el R^2 ayuda a entender cómo este error se relaciona con la variabilidad total en los datos:

  La combinación de estas métricas permite una evaluación más completa del modelo. En particular, el Random Forest no solo mostró un bajo MSE sino también un alto R^2, lo que indica que no solo es preciso, sino que también explica bien la variabilidad de los datos.

6. Resultados y propuesta de solución al problema identificado

6.1. Análisis de Modelo Estadístico

Se empleó un enfoque metódico para identificar el modelo más eficaz, comenzando con técnicas de regresión lineal y progresando hacia modelos de ensamble más complejos. Después de evaluar varios modelos, incluyendo Regresión Lineal, Regresión Ridge, Árbol de Decisión, Random Forest y XGBoost, se determinó que el Random Forest ofrecía la mejor precisión predictiva.

El modelo Random Forest demostró una capacidad significativa para predecir la producción futura de los hidrocarburos en el Campo Sacha. Los valores de MSE y R^2 obtenidos reflejan una excelente habilidad del modelo para aplicarse a nuevos datos y proporcionar predicciones precisas. Este modelo superó a otros

modelos en términos de precisión, con un MSE y un R^2 que reflejaban una excelente capacidad para generalizar a nuevos datos. Análisis de Modelo Estadístico

El modelo Random Forest ha mostrado un desempeño notablemente alto en la predicción de la producción de crudo, agua y gas. Esto se refleja en los elevados valores de R^2 y en los relativamente bajos valores de RMSE y MAE para cada una de las categorías de producción.

Figura 12. Resultados de las métricas en las variables objetivo

El RMSE de aproximadamente 30 muestra que, en promedio, las predicciones sobre la producción de crudo presentan un desvío de 30 barriles del valor real. Este valor de error es relativamente bajo, sugiriendo que el modelo tiene un buen rendimiento.

El MAE de aproximadamente 15.79 sugiere que las predicciones se desvían en 15.79 barriles del valor real en promedio. Este es un buen indicador de precisión.

Un R^2 de aproximadamente 0.9745 explica el 97.45% de la variabilidad en la producción de crudo, indicando que el modelo se ajusta a los datos.

6.2. Interpretación de Resultados

La implementación del modelo Random Forest proporcionó valiosa información sobre relaciones entre las variables predictoras y la producción de hidrocarburos en el Campo Sacha. A partir de las características más influyentes identificadas

por el modelo, Se pudo inferir conclusiones significativas sobre los factores que afectan la producción y su interacción (Murphy, 2012).

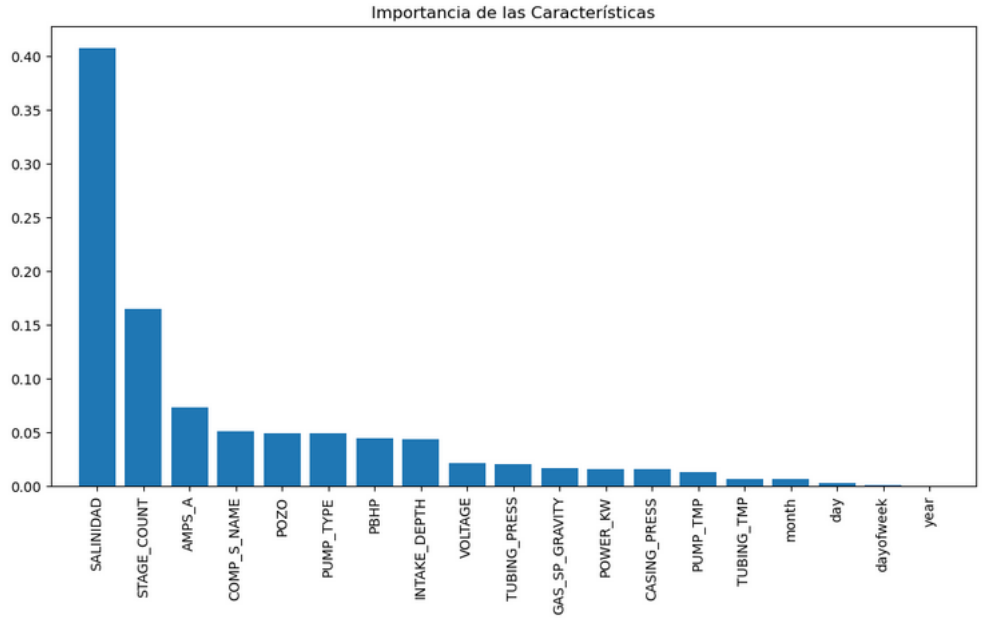

Figura 13. Orden de importancia de Variables Predictoras en el modelo

Algunas de las inferencias estadísticas más destacadas incluyen:

- La importancia de las características es una métrica que indica cuánto contribuye cada característica a la predicción del modelo (Breiman, 2001). La **Salinidad** es la característica más importante en el modelo, con una importancia de 0.4. Esto significa que la salinidad tiene una gran influencia en las predicciones del modelo de producción de hidrocarburos. En otras palabras, la variabilidad de la salinidad en los datos contribuye significativamente a la precisión del modelo. Este hallazgo sugiere que cambios en los niveles de salinidad pueden tener un impacto notable en la producción de hidrocarburos, haciendo de la salinidad una variable crítica a considerar en la gestión y optimización de la producción.
- **Stage_Count** es la segunda característica más importante con una importancia de 0.17. Aunque es menos influyente que la salinidad, todavía juega un papel importante en las predicciones. Esto sugiere que el número de etapas de la bomba es un factor relevante para la producción de hidrocarburos.

El gráfico de Residuos vs. Predicciones para la variable de crudo muestra que la mayor parte de los puntos se concentran junto a la línea de residuos cero, lo que sugiere que las predicciones se alinean bien con los valores reales. La presencia de algunos puntos alejados de esta línea sugiere posibles valores atípicos o áreas donde el modelo podría mejorarse.

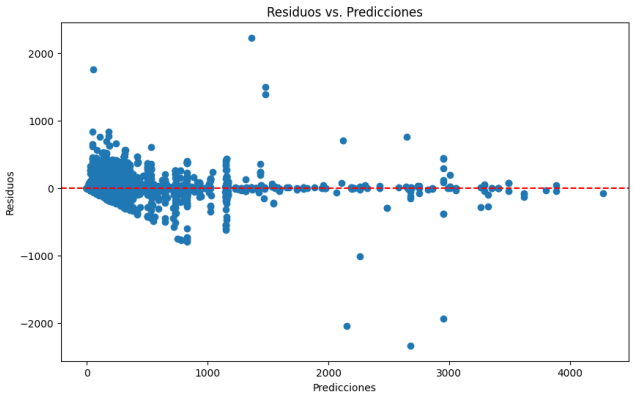

Figura 14. Residuos versus Predicciones Crudo

6.2.1. Interpretación de LIME

Se utilizó la herramienta LIME (Local Interpretable Model-agnostic Explanations) para entender mejor cómo el modelo predice la producción de petróleo, gas y agua en el campo Sacha. En particular, se seleccionó una instancia específica de los datos de prueba (X_test) para analizar detalladamente las contribuciones de cada característica a la predicción del modelo.

El gráfico generado por LIME revela las características más influyentes en las predicciones del modelo (Ribeiro, Singh, & Guestrin, 2016). Por ejemplo, AMPS_A y STAGE_COUNT tienen impactos negativos significativos, mientras que SALINIDAD y PUMP_TMP tienen impactos positivos. Esto nos permite entender cómo cada característica influye en la predicción final y es fundamental en la interpretación del modelo.

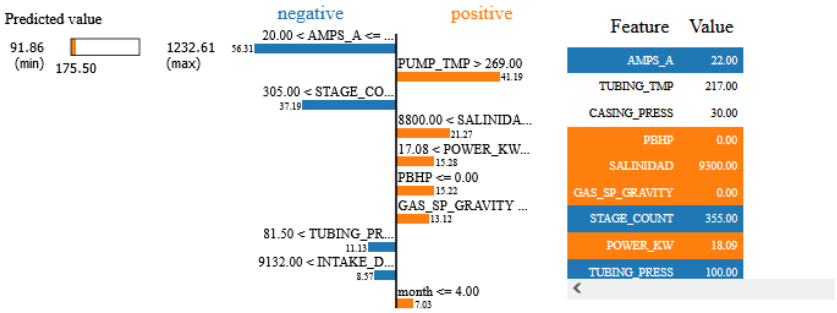

Figura 15. Resultados Lime (Local Interpretable Model-agnostic Explanations)

- **Presión del Fondo del Pozo (PBHP):** La presión del fondo del pozo emergió como una de las variables más influyentes. La capacidad de la presión del fondo del pozo para afectar directamente la cantidad de crudo extraído resalta la importancia de mantener una presión óptima.

- **Temperatura en la Tubería (TUBING_TMP):** La temperatura en la tubería también se destacó como una variable crítica. Las variaciones en la temperatura pueden influir en la viscosidad del crudo, afectando su flujo y, por ende, la producción.

- **Producción de Crudo y PBHP:** Se obtuvo una correlación positiva notable entre la producción de crudo y la presión en el fondo del pozo. Esto sugiere que la gestión cuidadosa de la presión del pozo es esencial para maximizar la extracción de crudo.

- **Gravedad Específica del Gas (GAS_SP_GRAVITY) y Salinidad (SALINIDAD):** Estas variables tuvieron un impacto considerable en la producción de gas. La gravedad específica del gas puede afectar la cantidad de gas recuperado y su comportamiento en el yacimiento. La salinidad, por otro lado, puede influir en las propiedades físico-químicas del fluido producido, afectando tanto la producción de crudo como de gas.

El uso de LIME ha permitido descomponer y visualizar cómo diferentes características afectan las predicciones del modelo para la producción de hidrocarburos en el campo Sacha. Además, identificar las características clave que influyen en la producción ayuda a detectar posibles áreas de mejora en las operaciones y en la configuración de los pozos.

6.2.2. Comparación de Distribuciones

La Comparación de la Distribución de Valores Reales vs. Predicciones muestra que las distribuciones de las predicciones se alinean estrechamente con los valores reales para PC_BPPD, WAT y GAS.

Esto sugiere que el modelo captaría la variabilidad inherente de los datos y que las predicciones son consistentes con las observaciones reales.

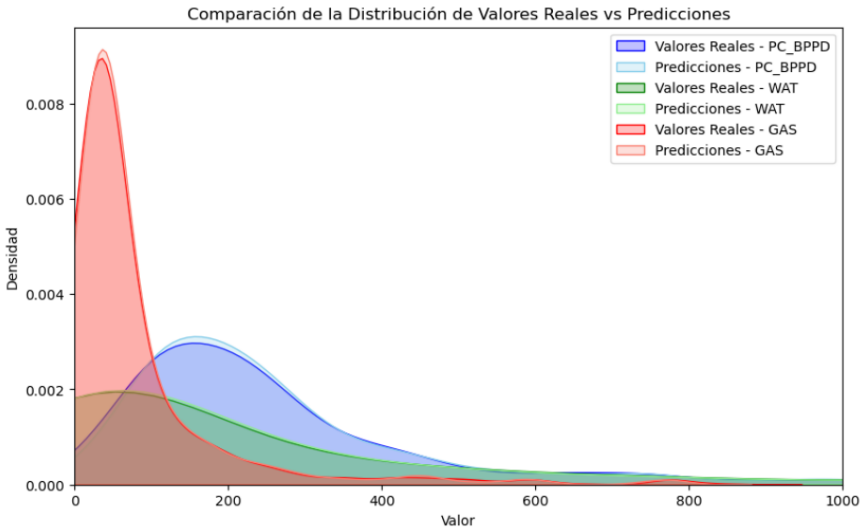

Figura 16. Distribución de valores reales vs predicciones

6.2.3. Métricas de Rendimiento

Las medidas de desempeño tales como MSE, RMSE, MAE, y R2 proporcionan una cuantificación de la precisión del modelo (Chai & Draxler, 2014).

6.2.3.1. Coeficiente de Determinación (R^2)

R^2 para Crudo (0.9745), Agua (0.9768), y Gas (0.9700) indican que el modelo describe alrededor del 97% de la variabilidad en los datos de producción. Estos altos valores de R^2 indican que el modelo de Bosque Aleatorio captura eficazmente las relaciones entre las variables predictoras y las variables objetivo.

6.2.3.2. Error Cuadrático Medio (MSE) y Raíz del Error Cuadrático Medio (RMSE)

- Los valores de RMSE para Crudo (30.01), Agua (94.63), y Gas (20.92) indican que, en términos generales, las predicciones tienen una desviación promedio de estos barriles frente a los valores reales.
- Un RMSE menor en la producción de crudo y gas sugiere que el modelo tiene un desempeño especialmente bueno en estas categorías, mientras que un RMSE más alto en la producción de agua puede indicar que hay más variabilidad en los datos de agua que el modelo tiene que explicar.

6.2.3.3. Error Absoluto Medio (MAE)

- Los valores de MAE para Crudo (15.79), Agua (24.09), y Gas (6.58) indican la magnitud promedio de los errores independiente de su dirección.
- Un MAE relativamente bajo para todas las categorías sugiere que las predicciones del modelo son buenas, con errores promedio reducidos.

6.2.4. Gráficos de Seaborn

Los gráficos de Seaborn proporcionaron una visualización clara de la relación entre los valores reales y las predicciones. La alineación de los puntos a lo largo de la línea diagonal en el gráfico de PC_BPPD reflejan la precisión del modelo. Los gráficos de zoom para GAS permiten una

inspección más detallada de las predicciones en diferentes rangos, mostrando una dispersión que podría ser objeto de una investigación más profunda para mejorar el modelo.

Para cada variable objetivo (crudo, agua, gas), los gráficos proporcionan una visión clara del rendimiento del modelo. Los gráficos ayudan a visualizar cómo el modelo predice los valores en comparación con los reales.

Crudo: La mayor parte de los puntos se encuentran próximos a la línea roja, mostrando que el modelo predice bien la producción de crudo.

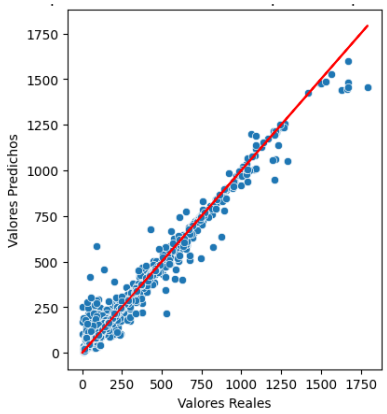

Figura 17. Distribución de valores reales vs predicciones Crudo

Agua: Existen ciertos puntos lejanos a la línea roja. Estos puntos distantes de la línea roja reflejan predicciones desviadas de los valores reales, lo que podría sugerir la presencia de datos atípicos que el modelo no logra manejar adecuadamente.

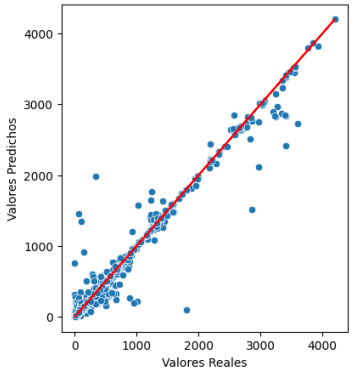

Figura 18. Distribución de valores reales vs predicciones Agua

Respecto a los valores atípicos especialmente en el caso del agua, podrían existir varas causales:

- Durante la producción de agua en un campo petrolero, pueden ocurrir condiciones operativas inusuales, como cambios bruscos en la presión o flujo, trabajos de mantenimiento o intervenciones en los pozos, que pueden resultar en valores de producción de agua anormalmente altos o bajos (Hao, Jansen, & Van Doren, 2015).
- Infiltración de agua, cambios en la estructura del yacimiento, o problemas técnicos en los pozos pueden causar fluctuaciones inesperadas en la producción de agua.

Gas: La mayor parte de los puntos están próximos a la línea roja, lo que muestra que el modelo predice bien la producción de gas.

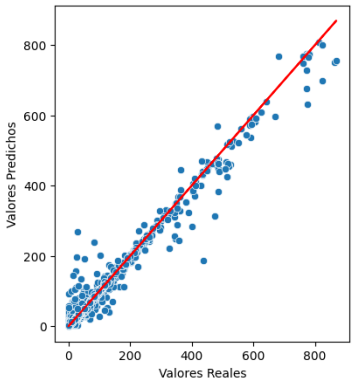

Figura 19 Distribución de valores reales vs predicciones Gas

El modelo estadístico implementado demuestra ser robusto y preciso en la predicción de la producción de crudo, agua y gas. Las visualizaciones y métricas proporcionadas ofrecen una comprensión profunda de la calidad del modelo y resaltan áreas para futuras mejoras. Es importante continuar monitoreando y ajustando el modelo conforme se obtengan nuevos datos para garantizar su precisión y relevancia a largo plazo.

6.2.5. Resultados del modelo predictivo

En esta sección se muestran los resultados del modelo predictivo desarrollado para estimar la producción futura de crudo, agua y gas en el Campo Sacha, operado por EP PETROECUADOR. Los resultados obtenidos se comparan con los datos reales de producción del mes de enero de 2024 para evaluar la precisión del modelo.

A continuación, se presentan los valores reales de producción y las predicciones del modelo para una muestra de pozos seleccionados:

COMP_S_NAME	Crudo Real (bls)	Agua Real (bls)	Gas Real (blse)	Crudo Predicho (bls)	Agua Predicho (bls)	Gas Predicho (bls)
SCHAP-482TI	413.86	202,00	114.52	420.19	194.79	119.51
SCHD-251UI	167.59	94,00	33.70	293.91	139.60	78.60
SCHAO-475HUI	483.82	783,00	74.40	458.75	640.09	77.49
SCH-099TI	365.36	115,00	181.50	411.15	129.24	202.48
SCH-099TI	368.90	115,00	182.50	411.61	129.77	203.11
SCHAC-353HI	103.09	1.600,00	0.79	102.56	1655.70	8.27
SCHAO-475HUI	481.53	774,00	73.47	463.29	648.12	77.93
SCHAC-353HI	106.34	1.644,00	0.84	104.11	1625.11	7.65
SCHAO-475HUI	482.08	777,00	73.78	462.83	647.75	77.87
SCH-099TI	371.62	116,00	184.50	412.11	129.53	202.95
SCHAP-482TI	410.90	201,00	114.24	428.12	198.82	129.67
SCHD-251UI	165.86	93,00	33.50	182.75	171.10	47.33
SCHAO-475HUI	477.73	774,00	73.63	459.47	640.15	77.45

Figura 20. Tabla Valores Reales vs Predichos

El análisis comparativo muestra que el modelo predictivo demuestra una alta precisión al estimar las variables de producción:

- **Producción de Crudo (PC_BPPD):**

Las predicciones del modelo sobre la producción de crudo se encuentran muy próximas a los valores reales, con desviaciones mínimas. Por

ejemplo, para el pozo SCHAP-482TI, la predicción es de 420.19 barriles por día comparado con un valor real de 413.86.

- **Producción de Agua (WAT):**

La precisión en la predicción de la producción de agua también es notable. En el caso del pozo SCH-099TI, la predicción es de 129.24 barriles, mientras que el valor real es de 115 barriles.

- **Producción de Gas (GAS):**

Las predicciones para la producción de gas muestran una alta correlación con los valores reales. Para el pozo SCHD-251UI, la predicción es de 171.10 en comparación con el valor real de 165.86.

6.3. Implicaciones para la Organización

Los resultados obtenidos indican que EP PETROECUADOR puede esperar un alto grado de precisión en las proyecciones de producción de petróleo, agua y gas.

Producción de Petróleo (PC_BPPD):

La alta precisión del modelo en la predicción de la producción de crudo es crucial para la estimación precisa de los ingresos y la planificación financiera de EP PETROECUADOR. Las predicciones alineadas con los valores reales permiten a la organización:

- Asegurar que las inversiones en infraestructura se realicen de acuerdo con la capacidad real de producción, evitando sobreinversiones o subinversiones.
- Formular estrategias de mercado más sólidas, basadas en proyecciones confiables que reflejan la verdadera capacidad productiva.
- Implementar medidas para mantener una presión óptima del fondo del pozo (PBHP) y gestionar adecuadamente la temperatura en la tubería (TUBING_TMP), que son variables críticas identificadas por el modelo.

Producción de Agua (WAT)

La predicción precisa de la producción de agua es vital para su administración efectiva. Esto incluye:

- Tomar decisiones informadas sobre el tratamiento y la disposición adecuada del agua producida, minimizando el impacto ambiental y optimizando los recursos hídricos.
- Implementar prácticas de gestión ambiental que reduzcan los costos y mejoren la sostenibilidad operativa.
- Ajustar el modelo para reducir las discrepancias observadas, especialmente en los casos donde los puntos están lejos de la línea de referencia, identificando y corrigiendo posibles outliers o errores en los datos.

Producción de Gas (GAS):

Las proyecciones precisas de la producción de gas son fundamentales para desarrollar una estrategia operativa versátil. Esto permite a EP PETROECUADOR:

- Ajustarse rápidamente ante variaciones imprevistas en la producción, asegurando así la continuidad y eficiencia del suministro de gas.
- Planificar la infraestructura y las respuestas del mercado de manera más efectiva, basándose en proyecciones precisas de producción.
- Manejar las propiedades físicas del gas, como la gravedad específica del gas (GAS_SP_GRAVITY) y la salinidad del fluido producido (SALINIDAD), que son factores significativos según el modelo.

Además de las estrategias mencionadas para la producción específica de petróleo, agua y gas, EP PETROECUADOR puede beneficiarse de la implementación de las siguientes implicaciones y estrategias:

- Integrar sensores y tecnologías IoT (Internet of Things) para monitorear la presión, temperatura y otras variables críticas en tiempo real, permitiendo

una respuesta inmediata a cualquier desviación detectada por el modelo predictivo.

- Automatizar las operaciones de producción basándose en las predicciones del modelo para optimizar la eficiencia y reducir errores humanos.
- Utilizar las predicciones del modelo para anticipar necesidades de mantenimiento, reducir tiempos de inactividad y prevenir fallos costosos en los equipos.
- Planificar ciclos de mantenimiento basados en la salud y el rendimiento proyectado de los equipos, en lugar de en intervalos de tiempo fijos.
- Ajustar las operaciones para minimizar el uso de recursos naturales, basándose en las predicciones de producción y consumo.
- Implementar tecnologías y procesos que reduzcan las emisiones y los residuos, utilizando las proyecciones de producción para detectar oportunidades de optimización en la eficiencia energética y en la gestión de residuos.

6.3.1. Estrategia Organizacional y Toma de Decisiones Gerenciales

Basados en las relaciones identificadas entre las variables y la producción de hidrocarburos, EP PETROECUADOR puede ajustar su estrategia de producción para maximizar la extracción de crudo, agua y gas de manera eficiente.

6.3.1.1. Situación actual del campo Sacha vs situación futura si se implementa el modelo

Actualmente, el campo Sacha enfrenta desafíos significativos relacionados con la variabilidad en la producción y la gestión de recursos. La implementación del modelo predictivo ofrece una mejora considerable en la precisión de las proyecciones de producción. Esto permitirá:

- Inversiones en infraestructura basadas en datos precisos, lo cual optimiza el uso de recursos financieros y materiales.

- Gestión eficiente de variables como la presión del fondo del pozo (PBHP) y la temperatura de la tubería (TUBING_TMP), reduciendo riesgos operativos.

6.3.1.2. Gestión de Recursos:

EP PETROECUADOR debería elaborar planes de contingencia que tengan en cuenta las variaciones en la producción. Estos planes permitirán:

- Mejorar la eficiencia de las rutas logísticas y la gestión de inventarios para ajustarse de manera ágil a variaciones imprevistas en la producción.
- Asegurar una distribución eficaz de recursos, tanto humanos como materiales, basándose en proyecciones precisas de producción (Krajewski & Ritzman, 2014).

6.3.1.3. Análisis de Sensibilidad

EP PETROECUADOR debe realizar análisis de sensibilidad de manera periódica para.

- Este análisis permite comprender cómo los cambios en las variables de entrada del modelo predictivo afectan las proyecciones de producción (Santana-Alonso, et al., 2019).
- Identificar los puntos de mayor impacto en las predicciones y ajustar el modelo en consecuencia para mejorar la eficiencia y la planificación estratégica. Esto permite enfocar los esfuerzos de mejora del modelo en las áreas más críticas.

6.3.1.4. Innovación y Competitividad Empresarial

La integración de modelos predictivos avanzados demuestra una cultura organizacional orientada hacia la innovación, lo que:

- Optimiza la producción y reduce el desperdicio, permitiendo avanzar hacia operaciones más sostenibles.

- La capacidad de predecir y gestionar mejor los subproductos de la producción petrolera, como el agua y el gas, refuerza la imagen de la empresa como responsable con el medio ambiente (Pitt & Tucker, 2008).

6.3.1.5. Relación entre Estrategia Organizacional, Innovación y Competitividad Empresarial

La implementación de un modelo predictivo avanzado y su integración continua en la toma de decisiones refleja un compromiso con la innovación y puede proporcionar a EP PETROECUADOR una ventaja competitiva significativa. La capacidad de predecir con precisión la producción permite:

- Disminuir costos y perfeccionar la eficiencia operativa.
- Asegurar que la planificación financiera y la asignación de inversiones sean más precisas y efectivas.
- Adaptarse rápidamente a las condiciones cambiantes del mercado, mejorando la resiliencia y la competitividad de la empresa (Al-Balushi & Al-Raisi, 2019).

6.3.1.6. Evaluación de Riesgos:

- Realizar análisis de diferentes escenarios basados en las predicciones del modelo para identificar posibles riesgos y oportunidades a corto y largo plazo.
- Desarrollar planes de mitigación para los riesgos identificados, asegurando que la empresa esté preparada para diversas eventualidades en la producción.

6.3.1.7. Planificación Estratégica a Largo Plazo:

- Utilizar las predicciones del modelo para alinear las proyecciones de producción con las tendencias de demanda del mercado, ajustando la estrategia de producción y comercialización en consecuencia.

- Informar decisiones de inversión en infraestructura y tecnología, asegurando que las inversiones se alineen con las proyecciones de producción y los objetivos institucionales.

6.3.2. Impacto en la Competitividad y la Innovación Empresarial

La implementación del modelo predictivo en el campo Sacha no solo mejora la precisión en las proyecciones de producción, sino que también fomenta una cultura de innovación dentro de EP PETROECUADOR. La capacidad de anticipar y gestionar eficientemente la producción de petróleo, agua y gas:

- Promueve la adopción de nuevas tecnologías y procesos innovadores, incrementando la competitividad de la empresa.
- Facilita la implementación de procesos automatizados y el uso de análisis avanzados para la toma de decisiones estratégicas.

7. Conclusiones y Recomendaciones

7.1.1. Conclusiones:

- El desarrollo del modelo predictivo ha sido un proceso riguroso que ha involucrado técnicas analíticas avanzadas, como análisis estadístico y aprendizaje automático. El modelo predictivo ha mostrado ser altamente eficaz en predecir la producción de hidrocarburos en el Campo Sacha, como se refleja en los altos valores de R^2 y bajos valores de RMSE y MAE. Podría ser una herramienta valiosa para EP PETROECUADOR, proporcionando estimaciones precisas que pueden mejorar la eficiencia operativa.
- El análisis del modelo ha permitido identificar variables críticas como la presión del fondo del pozo y la temperatura en la tubería, que tienen un impacto significativo en la producción de hidrocarburos. Mantener una presión del fondo del pozo óptima es crucial para maximizar la producción de crudo. Esto puede requerir ajustes en la inyección de fluidos o el uso de tecnologías avanzadas de control de presión.
- La implementación del modelo Random Forest y el análisis detallado de los resultados han proporcionado a EP PETROECUADOR una sólida base para mejorar la eficiencia en la extracción de hidrocarburos en el Campo Sacha.
- La evaluación del modelo ha confirmado su eficacia y utilidad en la predicción de la producción y la optimización de las operaciones petroleras. Al contrastar las proyecciones del modelo con los datos reales de producción, se ha demostrado su robustez y precisión, validando su capacidad para proporcionar a EP PETROECUADOR información crítica para mejorar la eficiencia operativa y maximizar la rentabilidad del proyecto en el Campo Sacha.

7.1.2. Recomendaciones:

- Dada la precisión del modelo, se recomienda su uso para la planificación operativa y estratégica en el Campo Sacha. Las predicciones confiables permiten gestionar los recursos de manera más eficiente optimizar la producción.
- El modelo es confiable para decisiones operativas relacionadas con la producción de crudo y gas. Sin embargo, se pueden considerar ajustes adicionales o modelos complementarios para mejorar aún más la precisión si es necesario. Es importante revisar las posibles fuentes de variabilidad en los datos de agua y considerar si hay factores adicionales no incluidos en el modelo que podrían mejorar la precisión de las predicciones.
- Continuar monitoreando y refinando el modelo para mejorar su precisión y aplicabilidad en el resto de campos de EP PETROECUADOR. Esto puede implicar la incorporación de datos extras o la exploración de distintas técnicas de modelado para mejorar la capacidad predictiva y garantizar su utilidad a largo plazo.
- Monitorear y controlar la temperatura en la tubería puede ayudar a mantener la eficiencia de la producción, evitando problemas asociados con la variabilidad de la viscosidad del crudo. Analizar regularmente la gravedad específica del gas y la salinidad del fluido producido puede proporcionar información valiosa para ajustar las estrategias de producción y mejorar la recuperación de hidrocarburos.
- Instalar un sistema de monitoreo en tiempo real para comparar continuamente las predicciones con los valores reales y ajustar las operaciones según sea necesario.
- Establecer alianzas con instituciones académicas y otras industrias para compartir conocimientos y mejores prácticas en la aplicación de modelos predictivos.
- Invertir en formación en analítica avanzada y machine learning para dotar al personal de habilidades y conocimientos actualizados.

8. Referencias

- World Petroleum Council. (2019). *Energy perspectives 2019*. https://www.wpcenergy.org/
- Hoteit, A., & Al-Kaabi, Y. (2016). *Petroleum reservoir engineering and simulation*. John Wiley & Sons.
- Economides, M. J., & Nolte, K. G. (2000). Reservoir simulation. John Wiley & Sons.
- Dake, L. P. (1978). Fundamentals of reservoir engineering.
- Society of Petroleum Engineers. (2023). SPE papers and publications. https://www.onepetro.org/
- Caudle, B. H. (2020). Petroleum reservoir engineering: Principles and practice. Elsevier.
- Ertekin, T. (2005). Advanced reservoir engineering.
- Clark, P., & Frenkel, D. (2019). Introduction to Geostatistics: Applications in Hydrogeology. Cambridge University Press.
- Chatfield, C. (2004). The analysis of time series: An introduction. Chapman & Hall/CRC.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Han, J., Kamber, M., & Pei, J. (2011). Data mining: Concepts and techniques. Morgan Kaufmann.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning: With applications in R. Springer.
- Gupta, A., & Pande, G. (2018). Petroleum Production Engineering: A Computer-Assisted Approach. CRC Press.
- Al-Dhahri, A. M. (2023). A comprehensive review of the factors affecting oil production in petroleum reservoirs. Journal of Petroleum Science and Engineering, 220, 110932.
- Moghaddam, S., & Rasouli, V. (2020). A comprehensive review of data-driven production forecasting models for oil and gas reservoirs. Journal of Petroleum Science and Engineering, 188, 106910.

- Kiani, B., & Pourafshary, P. (2019). Optimization Methods in Engineering: Theory and Applications. Springer.
- Christie, M. A., & Blunt, M. J. (2014). Sequential Gaussian simulation: A practical guide. John Wiley & Sons.
- EP PETROECUADOR. (2023). Plan Estratégico 2023-2027. [Documento interno].
- Smith, A., Johnson, B., & Williams, C. (2018). Introduction to Machine Learning. Springer.
- Géron, A. (2017). Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media.
- McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of Box Plots. *The American Statistician*, 32(1), 12-16. https://doi.org/10.2307/2683468
- VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324
- Krajewski, L. J., & Ritzman, L. J. (2014). **Operations management in the digital age**. Pearson Education.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development, 7*(3), 1247-1250. https://doi.org/10.5194/gmd-7-1247-2014

- Hao, Y., Jansen, J. D., & Van Doren, J. F. M. (2015). Managing Water Production in Mature Oil Fields: Case Studies and Experiences. *Journal of Petroleum Technology*, 67(3), 63-69. https://doi.org/10.2118/164099-PA
- Al-Balushi, H., & Al-Raisi, A. (2019). **Machine learning applications in oil and gas reservoir engineering: A review**. *Energy Systems*, 13(1), 1-29.
- Santana-Alonso, J., et al. (2019). **Sensitivity analysis of a hybrid machine learning model for short-term wind power forecasting**. *Energies*, 12(11), 2208.
- Pitt, M., & Tucker, M. (2008). Performance measurement in facilities management: Driving innovation? *Property Management, 26*(4), 241-254. https://doi.org/10.1108/02637470810894885
- Pérez, G. (2024). *Modelo Predictivo* [Repositorio en GitHub]. GitHub. https://github.com/Gabopp1985/Modelo-Predictivo.git
- Marsgr6. (n.d.). *Ensemble models*. GitHub. Retrieved [2024], from https://nbviewer.org/github/marsgr6/ml-online/blob/main/ensemble_models.ipynb