ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.1033

Data-driven selection of artificial lift systems using machine-learning algorithms: Lago Agrio field case study

Ing. Jean Pierre-Mendia Cadena[1], MSc. Freddy Paúl-Carrión Maldonado[1], MSc. Jorge Rodrigo-Lliguizaca Dávila[2]
[1]Litoral Polytechnic Higher School, Ecuador, *jmendia@espol.edu.ec, fpcarrio@espol.edu.ec, jorollig@espol.edu.ec*
[2]*University of Bergen, Norway, jolli5902@uin.no*

This study presents a data-driven workflow to optimize artificial lift system (ALS) selection for wells in the Oriente Basin, Ecuador. The goal is to support engineers in choosing the most suitable ALS based on production and operational characteristics. Proper ALS selection is critical to maintain stable production, reduce unnecessary energy use, and minimize failures caused by mismatches between reservoir conditions and lift mechanisms. Historical well and production data were compiled and processed through data cleaning, feature engineering, and class-balancing techniques to improve representation of underused ALS categories. Multiple multiclass machine-learning classifiers were trained to predict the recommended ALS using key operational parameters. The best model was embedded in a web-based application that allows users to input well data and obtain data-driven ALS recommendations. Among the evaluated algorithms, XGBoost and a Stacking ensemble achieved the strongest performance, with test accuracies above 99%, while Random Forest and Decision Tree models reached about 96%. Overall evaluation shows high predictive capability. Comparison with field installations yielded an agreement of 83.3% between model recommendations and deployed ALSs. Although field choices do not always match model outputs, results indicate that the models provide robust and consistent predictions under current data conditions. The proposed workflow demonstrates that machine-learning–based ALS selection is a practical and reliable decision-support approach for wells with similar characteristics. Its use can help standardize selection criteria, reduce operational uncertainty, and improve production management efficiency across assets.

Keywords: Machine learning, Data science, Artificial lift systems, Hydrocarbon production

Selección de sistemas de levantamiento artificial mediante clasificación con aprendizaje automático: Caso de estudio Campo Lago Agrio

Ing. Jean Pierre-Mendia Cadena[1], MSc. Freddy Paúl-Carrión Maldonado[1], MSc. Jorge Rodrigo-Lliguizaca Dávila[2]
[1]Escuela Superior Politécnica del Litoral, Ecuador, *jmendia@espol.edu.ec, fpcarrio@espol.edu.ec, jorollig@espol.edu.ec*
[2]*University of Bergen, Norway, jolli5902@uin.no*

Este estudio presentó un flujo de trabajo basado en datos para la selección de sistemas de levantamiento artificial (SLA) en pozos de la Cuenca Oriente, Ecuador. El objetivo fue apoyar a los ingenieros en la identificación del SLA más adecuado para un pozo, considerando sus características de producción y operación. La selección apropiada de un SLA fue esencial porque permitió mantener una producción estable, reducir el consumo innecesario de energía y minimizar fallas derivadas de desajustes entre las condiciones del yacimiento y el mecanismo de levantamiento seleccionado. Para ello, se recopilaron datos históricos de producción y se procesaron mediante procedimientos de limpieza, ingeniería de características y de balanceo de clases, con el fin de mejorar la representatividad de los tipos de SLA con menor presencia. Posteriormente, se entrenó un conjunto de modelos de clasificación multiclase de aprendizaje automático para predecir el SLA óptimo a partir de parámetros operativos clave. El modelo con mejor desempeño se integró en una aplicación web que permitió al usuario ingresar información del pozo y obtener recomendaciones de selección basadas en datos. Entre los algoritmos evaluados, XGBoost y el ensamble Stacking presentaron el mayor desempeño, alcanzando exactitudes superiores al 99% en el conjunto de prueba. En contraste, Random Forest y Decision Tree mostraron valores cercanos al 96%. La evaluación evidenció una alta capacidad predictiva. Se observó una concordancia del 83.3% entre las recomendaciones del modelo basado en datos y los SLA instalados en campo. Estos resultados mostraron que, aunque las decisiones de campo no siempre coincidieron con las recomendaciones del modelo, los algoritmos demostraron un potencial predictivo robusto bajo las condiciones de datos disponibles. Por lo tanto, la selección de SLA basada en aprendizaje automático se consideró un método confiable y práctico para apoyar decisiones de ingeniería en pozos con condiciones operativas similares. Su implementación contribuyó a estandarizar criterios de selección, reducir la incertidumbre operativa y mejorar la eficiencia de la gestión de producción en distintos activos.

Palabras clave: Aprendizaje automático, Ciencia de datos, Sistemas de levantamiento artificial, Producción de hidrocarburos.

I. Introducción

La producción de hidrocarburos en campos maduros enfrenta desafíos técnicos y operativos, entre ellos la declinación natural de la presión del yacimiento, el incremento del corte de agua y la complejidad en el manejo de fluidos [1], [2]. En este contexto, los sistemas de levantamiento artificial (SLA) resultan indispensables para sostener tasas de producción cuando la energía natural del yacimiento es insuficiente [3]. Sin embargo, la selección del SLA sigue siendo un punto crítico de decisión: una elección inadecuada puede traducirse en ineficiencias operativas, mayor frecuencia de intervenciones, incremento de los costos de mantenimiento y pérdidas de producción [4].

Tradicionalmente, la selección de SLA se ha sustentado en criterios empíricos, metodologías basadas en índices de desempeño, reglas heurísticas y, principalmente, en la experiencia del ingeniero de producción [5], [6]. Aunque estos métodos han sido ampliamente utilizados en la práctica ingenieril, su capacidad para capturar de manera integral la interacción no lineal entre múltiples variables operativas y productivas es limitada, lo que puede derivar en decisiones poco consistentes o ineficientes, especialmente en campos maduros con condiciones dinámicas de operación [7], [8]. En consecuencia, el proceso de selección tradicional de sistemas de levantamiento artificial tiende a carecer de reproducibilidad objetiva y a depender ampliamente del criterio individual del ingeniero, lo que puede traducirse en decisiones inconsistentes y con menor robustez frente a la incertidumbre operativa y la variabilidad de las condiciones de pozo.

En los últimos años, la ciencia de datos y las técnicas de aprendizaje automático han demostrado un alto potencial para apoyar la toma de decisiones en la industria petrolera [9] El análisis de grandes volúmenes de datos históricos permite identificar patrones ocultos, relaciones no lineales y comportamientos complejos que difícilmente pueden capturarse mediante métodos convencionales [10]. En particular, los modelos de clasificación basados en aprendizaje automático ofrecen una alternativa prometedora para la selección de sistemas de levantamiento artificial, al permitir evaluar simultáneamente múltiples variables operativas y productivas [11].

Bajo este enfoque, el presente trabajo propone un modelo de apoyo a la decisión basado en el aprendizaje automático para la selección de sistemas de levantamiento artificial en pozos petroleros del campo Lago Agrio, ubicado en la cuenca Oriente del Ecuador. El estudio se fundamenta en el procesamiento y análisis de datos históricos reales de producción y operación, a partir de los cuales se entrenan y evalúan distintos modelos de

clasificación multiclase orientados a establecer una relación sistemática entre las condiciones operativas del pozo y el sistema de levantamiento más adecuado [12]. La contribución del artículo consiste en la evaluación comparativa del desempeño de los modelos propuestos y en una validación inicial frente a sistemas de levantamiento artificial previamente instalados en campo bajo criterios tradicionales, lo que permite analizar su consistencia y resalta el potencial de esta metodología como una herramienta complementaria, objetiva y reproducible para la toma de decisiones en ingeniería de producción.

II. Metodología

La fig. 1 presenta la metodología basada en ML para la selección de SLA; en ella se detalla cómo se organizó en un flujo de trabajo basado en datos que abarcó desde la adquisición de datos hasta la implementación de un aplicativo web para apoyar la selección de un sistema de levantamiento artificial, incluyendo análisis exploratorio, preprocesamiento, ingeniería y selección de características, entrenamiento y evaluación de los modelos, e integración del modelo seleccionado en el aplicativo web.

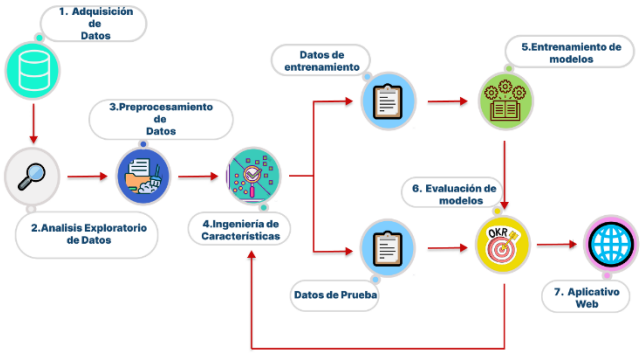

Fig. 1 Flujo de trabajo de la metodología de la selección de SLA utilizando algoritmos de ML.

A. *Fuentes de datos y variables*

Los datos recopilados provienen de registros históricos de producción y operación de pozos del campo Lago Agrio (cuenca Oriente, Ecuador), proporcionados por la compañía EP Petroecuador. En este trabajo se consideraron los siguientes sistemas de levantamiento artificial: bombeo hidráulico (BH), bombeo electrosumergible (BES) y bombeo mecánico (BM). Además, la base de datos incluyó parámetros como las tasas de producción de fluidos, las propiedades de los fluidos y la profundidad, así como el tipo de sistema de levantamiento artificial (SLA) como variable objetivo.

En la Tabla I se resumen las variables empleadas en la construcción del modelo de datos, sus unidades, su descripción y su rol en el enfoque propuesto. Para facilitar su interpretación, las variables se organizaron según el tipo de información que aportan. En primer lugar, las variables de producción (BFPD y BPPD) describen tasas de producción de fluido total y de petróleo. En segundo lugar, las variables asociadas a propiedades y composición del fluido (BSW, API y GOR), parámetros que influyen en el comportamiento de flujo y en el desempeño de los sistemas de levantamiento. Adicionalmente, las variables de presión (Pyac y Psep) capturan la energía disponible en el yacimiento y las condiciones de superficie en el separador, factores que condicionan la necesidad y viabilidad de cada alternativa de levantamiento [3]. Por su parte, las variables de profundidad (TVD_min y TVD_max) describen el intervalo de profundidad verdadera de la arena productora, información que se relaciona con la zona de aporte de fluidos y condiciona la factibilidad y el desempeño de ciertas alternativas de levantamiento. La variable de Fecha se incluyó para mantener la referencia temporal de los registros y permitir análisis de tendencias. Finalmente, SLA se definió como la variable objetivo, mientras que el resto se empleó como variables predictoras

TABLA I
VARIABLES INICIALES PARA LA SELECCIÓN EN EL MODELADO

Variables	Unidad	Descripción	Tipo
BFPD	Bbl/d	Producción total de fluido	Predictora
BPPD	Bbl/d	Producción de petróleo	Predictora
BSW	%	Corte de agua y sedimentos	Predictora
API	°API	Gravedad API del crudo	Predictora
GOR	Scf/bbl	Relación gas-petróleo	Predictora
Pyac	Psia	Presión del yacimiento registrado	Predictora
Psep	Psia	Presión del separador registrado	Predictora
Fecha	dd/mm/aaaa	Fecha del registro	Predictora
TVD_max	Ft	Profundidad máxima verdadera total	Predictora
TVD_min	Ft	Profundidad mínima verdadera total	Predictora
SLA	Categórica	Sistema de levantamiento artificial	Objetivo

B. *Análisis exploratorio de datos (EDA)*

Se realizó un análisis exploratorio de datos (EDA) con el fin de caracterizar el conjunto de datos y entender el comportamiento general de las variables consideradas. Para ello, se calcularon estadísticos descriptivos (media, mínimo, máximo y varianza) y se revisó la distribución de las variables productivas (BFPD, BPPD), propiedades del fluido (BSW, API, GOR) y de profundidad. En particular, a partir de los

límites de profundidad del intervalo productor (TVD_min y TVD_max), previa conversión a formato numérico para su análisis. Asimismo, se evaluó la distribución del sistema de levantamiento artificial como variable objetivo, identificándose un desbalance entre las clases, lo cual justificó la aplicación posterior de un procedimiento de balanceo para reducir el sesgo durante el entrenamiento.

C. Preprocesamiento de datos

Con la finalidad de garantizar la consistencia y la calidad del conjunto de entrenamiento, se aplicó un flujo de preprocesamiento de datos que incluyó (i) eliminación de registros duplicados, (ii) tratamiento de valores faltantes mediante imputación con mediana en variables con ausencias parciales y (iii) corrección de tipos de datos para asegurar el manejo adecuado de categorías[10]. Adicionalmente, se mitigó el efecto de los valores extremos mediante winsorización, conservando el tamaño muestral sin que los valores atípicos dominaran el entrenamiento [13]. Finalmente, las variables numéricas se escalaron mediante normalización mínimo-máximo para homogeneizar las magnitudes y estabilizar el entrenamiento de los algoritmos [10].

Debido a que la variable objetivo presentó un desbalance entre clases, se aplicó un procedimiento de balanceo de clases. El conjunto inicial presentó un desbalance entre los tipos de SLA, con una clase minoritaria de baja representatividad. Para reducir el sesgo en la clasificación multiclase, se aplicó un balanceo de clases mediante simulación paramétrica tipo Monte Carlo sobre la clase minoritaria, ajustando las distribuciones a partir de estadísticos descriptivos (media y desviación estándar) y generando registros sintéticos bajo restricciones físicas, como caudales no negativos. Este procedimiento mejoró la representatividad de la clase minoritaria y estabilizó la capacidad de generalización de los modelos [14].

D. Ingeniería y selección de características

Se construyeron variables derivadas para capturar información operativa relevante. En particular, el intervalo productor, representado por TVD_min y TVD_max, se encontraba originalmente en formato texto; por ello, se extrajeron ambos límites, se transformaron a formato numérico y se calculó TVD_mean como la profundidad verdadera media del intervalo productor. Debido a que TVD_mean resume la misma información y reduce la redundancia, TVD_min y TVD_max no se incluyeron como variables predictoras. La variable Fecha se mantuvo únicamente para organización temporal y análisis de los registros, sin incorporarse al modelado. Por último, variables como Pyac y Psep fueron descartadas debido a que esta información no estaba completa para todos los pozos. Posteriormente, se evaluaron asociaciones lineales entre variables numéricas mediante la correlación de Pearson, a partir de una matriz de correlación, con la finalidad de identificar variables altamente correlacionadas y reducir la redundancia.

La Tabla II resume las variables finales seleccionadas, junto con sus unidades y descripción. Como variables predictoras para el modelado se seleccionaron BFPD, BPPD, BSW, API, GOR y TVD_mean, considerando su aporte informativo y su baja redundancia relativa, mientras que SLA se definió como la variable objetivo (clase multiclase) a predecir.

TABLA II
VARIABLES SELECCIONADAS PARA MODELOS

Variables	Unidad	Descripción	Tipo
BFPD	Bbl/d	Producción total de fluido	Predictora
BPPD	Bbl/d	Producción de petróleo	Predictora
BSW	%	Corte de agua y sedimentos	Predictora
API	°API	Gravedad API del crudo	Predictora
GOR	Scf/bbl	Relación gas-petróleo	Predictora
TVD_mean	Ft	Profundidad media verdadera total	Predictora
SLA	Categórica	Sistema de levantamiento artificial	Objetivo

E. Entrenamiento y evaluación de modelos

El problema se formuló como una clasificación multiclase para predecir el tipo de SLA. Se seleccionaron algoritmos orientados a datos tabulares y capaces de modelar relaciones no lineales con distintos niveles de complejidad, con la finalidad de comparar desempeño e interpretabilidad. En particular, el algoritmo Decision Tree se utilizó como modelo base por su facilidad de interpretación, Random Forest se usó por su capacidad de reducir varianza y mejorar estabilidad frente a un único árbol, XGBoost por su elevado desempeño en clasificación sobre variables heterogéneas mediante el boosting y regularización y por último Stacking Classifier para combinar modelos y aprovechar fortalezas complementarias, buscando mejorar la generalización [15], [16]. El conjunto se dividió en entrenamiento y prueba mediante una partición 80/20, preservando la proporción entre las clases. Los datos se dividieron en dos conjuntos: uno de entrenamiento, denominado “train”, y otro de validación, llamado “validation”.

El conjunto de datos se dividió en entrenamiento y prueba mediante una partición 80/20 estratificada, preservando la proporción entre clases. Posteriormente, el ajuste de hiperparámetros se realizó exclusivamente con el conjunto de entrenamiento, aplicando validación cruzada k-fold. En este proceso, la métrica de entrenamiento (train) y en validación (validation) se obtuvo promediando los resultados de los pliegues, lo que permitió identificar configuraciones con mejor capacidad de generalización y menor sobreajuste. Finalmente, con los parámetros seleccionados, el modelo se reentrenó con todo el conjunto de entrenamiento y se reportó como el desempeño final sobre el conjunto de prueba (test)

La evaluación de la generalización de los modelos se realizó utilizando métricas para problemas de aprendizaje supervisado de clasificación multiclase tales como: exactitud, precisión, sensibilidad y Puntuación F1, derivadas a partir de la matriz de confusión [17]. La exactitud cuantifica la proporción total de predicciones correctas, la precisión mide la confiabilidad de las predicciones positivas, la sensibilidad refleja la capacidad del modelo para identificar correctamente los casos positivos y el Puntuación F1 resume el balance entre la precisión y la sensibilidad. Para consolidar la estimación del desempeño y reducir el sobreajuste, se aplicó la validación cruzada durante el ajuste de hiperparámetros.

$$Exactitud = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

$$Precisión = \frac{TP}{(TP + FP)} \quad (2)$$

$$Sensibilidad = \frac{TP}{(TP + FN)} \quad (3)$$

$$F1\ Puntuación = 2 * \left(\frac{Precisión * Sensibilidad}{Precisión + Sensibilidad}\right) \quad (4)$$

F. Aplicativo Web

Como complemento a los modelos predictivos, se implementó un aplicativo web interactivo denominado OptiLift, cuyo propósito fue facilitar el uso práctico del modelo seleccionado para la predicción del sistema de levantamiento artificial a partir de variables operativas y productivas. El aplicativo se desarrolló utilizando el framework panel (Python), integrando una interfaz gráfica orientada a la interacción del usuario y la visualización de resultados.

La arquitectura del aplicativo se organizó en dos componentes: Backend en Python, apoyado en librerías para manejo y transformación de datos y generación de visualizaciones interactivas. El Frontend se construyó con la librería HoloViz para integrar la interfaz con los modelos entrenados y sus respectivas predicciones.

La interfaz gráfica incluyó módulos para: carga de datos (formato CSV/XLSX), visualización exploratoria mediante gráficos de distribución y series temporales, ingreso manual de variables independientes del pozo y ejecución del modelo para generar la recomendación del sistema de levantamiento artificial, presentando así, los resultados de manera tabular Mostrando la clase predicha y las probabilidades estimadas para cada alternativa de SLA, lo que permitió identificar no solo el sistema más probable, sino también las opciones con mayor nivel de recomendación en caso de desempeños similares.

III. ANÁLISIS Y RESULTADOS

A. Comparación cuantitativa del desempeño

Con el fin de analizar el comportamiento de los modelos durante el ajuste y reducir el riesgo de sobreajuste, se construyeron curvas de validación comparando el desempeño en entrenamiento y validación al variar un hiperparámetro representativo por modelo. En general, se observó un desempeño alto y estable para los métodos de ensamble, con una brecha reducida entre entrenamiento y validación, mientras que el Decision tree mostró mayor sensibilidad al parámetro de complejidad.

La Fig. 2 presenta la exactitud (accuracy) en función del hiperparámetro max_depth para el modelo de Decision Tree, comparando el desempeño de train y validation. Al incrementar los valores de max_depth desde valores bajos hasta un rango intermedio, la exactitud aumentó de forma marcada en ambos conjuntos, lo que evidenció una mejora en el modelo. A partir de este rango, el desempeño en validation tendió a estabilizarse, mientras el conjunto de train continuó incrementándose ligeramente. Por lo que se identificó una zona adecuada para la selección de la profundidad sin aumentar innecesariamente la complejidad del modelo.

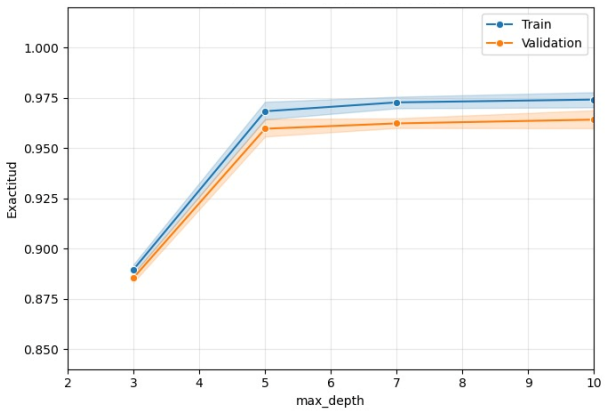

Fig. 2 Curva de validación del modelo Decision Tree en entrenamiento (train) y validación (validation) en función de la profundidad máxima (max_depth).

La Fig. 3 presenta la exactitud en función del hiperparámetro n_estimators para el modelo Random Forest comparando el desempeño de train y validation. En el rango evaluado (100-300 árboles), se observó, un comportamiento estable, con una leve mejora de desempeño en validation a medida que el número de estimadores aumentó. Asimismo, la diferencia entre train y validation se mantuvo reducida, lo que sugirió una generalización consistente y baja sensibilidad del modelo al parámetro evaluado. En consecuencia, incrementos adicionales en n_estimators aportaron mejoras marginales, por lo que la selección del valor final se sustentó en el equilibrio entre el desempeño y costo computacional.

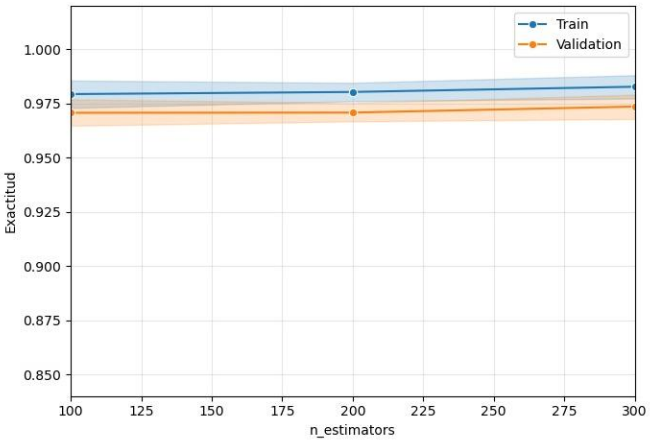

Fig. 3 Curva de validación del modelo Random Forest en entrenamiento (train) y validación (validation) en función del número de árboles (n_estimators)

En el caso del modelo entrenado con el algoritmo XGBoost, la Fig. 4 muestra una variación mínima del desempeño en el conjunto validation al modificar max_depth dentro del rango analizado, mientras que train se mantuvo cercano a valores máximos. Este patrón reflejó estabilidad del modelo respecto a profundidad evaluada. La elección de max_depth se orientó a mantener una complejidad moderada sin afectar el desempeño de la validación.

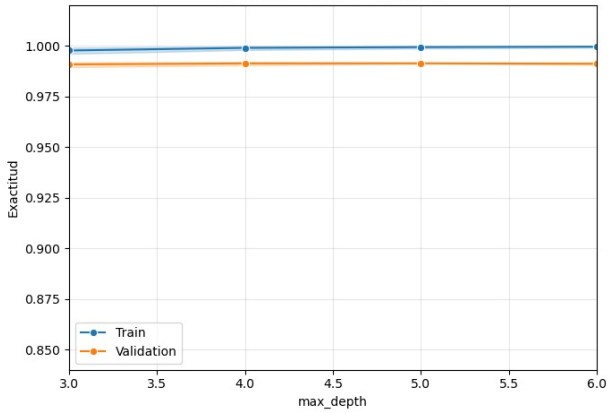

Fig. 4 Curva de validación de modelo XGBoost en entrenamiento (train) y validación (validation) en función de la profundidad máxima (max_depth)

Finalmente, la curva correspondiente al Stacking Classifier (Fig. 5) mantuvo valores elevados y prácticamente constantes en ambos conjuntos al varía el hiperparámetro final_estimator_n_neighbors. Esto indicó una baja sensibilidad del ensamble al parámetro evaluado en el intervalo considerado.

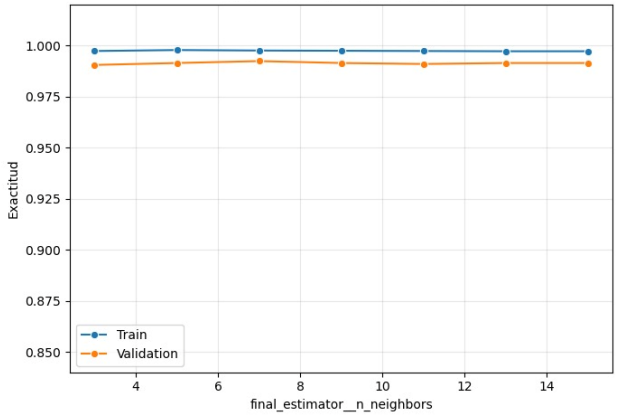

Fig. 5 Curva de validación del modelo Stacking Classifier en entrenamiento (train) y validación (validation) en función del número de vecinos del estimador final (final_estimator_n_neighbors)

Una vez seleccionados los hiperparámetros a partir del proceso de ajuste, se entrenaron los modelos finales y se evaluaron en el conjunto de prueba utilizando la exactitud, precisión sensibilidad y Puntuación F1. La Tabla III resume los resultados obtenidos para los cuatro algoritmos evaluados.

TABLA III
COMPARACIÓN DE MÉTRICAS DE DESEMPEÑO EN TEST PARA LOS MODELOS EVALUADOS

Modelo	Exactitud	Precisión	Sensibilidad	F1 Puntuación
Decision Tree	0.964	0.964	0.964	0.964
Random Forest	0.986	0.961	0.960	0.960
XGBoost	0.992	0.992	0.992	0.992
Stacking Classifier	0.987	0.991	0.991	0.991

En general, los modelos de ensamble presentaron el mejor rendimiento: XGBoost alcanzó la mayor exactitud (0.992) y mantuvo un patrón similar en las demás métricas, mientras que Stacking Classifier obtuvo un desempeño competitivo (exactitud 0.987). En contraste, Decision Tree presentó el rendimiento más bajo (0.964), evidenciando limitaciones para capturar patrones complejos en un escenario multiclase.

B. Análisis del modelo seleccionado

Aunque XGBoost presentó un desempeño marginalmente superior en test, ambos modelos mostraron un desempeño comparable en validación cruzada (CV), alcanzando el mismo valor máximo de CV. Además, la brecha entre el entrenamiento y la validación fue prácticamente equivalente, por lo que no se

observó una diferencia significativa en tendencia de sobreajuste. En este contexto, se seleccionó Stacking Classifier como modelo final por criterios de robustez del ensamble y facilidad para incorporar modelos base adicionales en futuras iteraciones, manteniendo un rendimiento competitivo.

Con el fin de sustentar la comparación durante el ajuste de hiperparámetros, la Tabla IV presenta el mejor desempeño promedio en validación cruzada (CV) y la brecha train–CV (gap) para XGBoost y Stacking. En ambos casos se obtuvo el mismo valor máximo de CV y un gap prácticamente equivalente, lo que sugiere un comportamiento similar en términos de generalización bajo el esquema de validación aplicado.

TABLA IV
DESEMPEÑO EN VALIDACIÓN CRUZADA Y BRECHA TRAIN–CV DE LOS MODELOS EVALUADOS

Modelo	Best CV (mean)	Gap (Train-CV)
XGBoost	0.992468	0.005295
Stacking Classifier	0.992468	0.005178

Con la finalidad de analizar el comportamiento del modelo seleccionado, se examinó la matriz de confusión en el conjunto de prueba (Fig.6). Esta matriz muestra un desempeño consistente con 526 clasificaciones correctas de 531 registros, el nivel de acierto fue cercano al 99 % y únicamente 5 errores. Para la clase BES, el modelo identificó correctamente 164 de 165 casos, presentando confusión hacia BH y ninguna hacia BM. Para la clase BH se observó un comportamiento perfecto con 215 de 215 casos. En la clase BM, el modelo clasificó correctamente 147 de 151 registros, con 4 casos erróneos hacia BH.

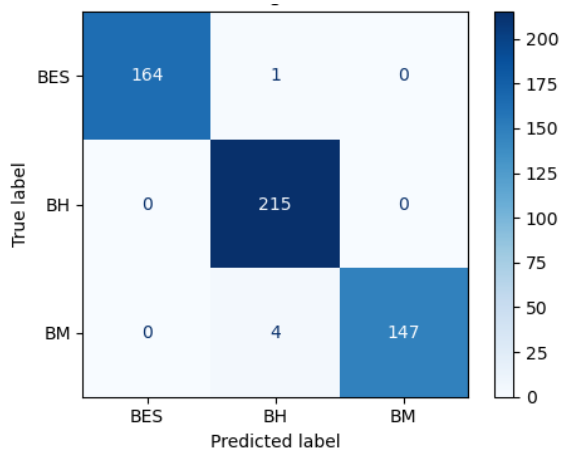

Fig. 6 Matriz de confusión del Stacking Classifier en el conjunto de prueba ("Test"), mostrando las etiquetas reales ("True label") y predichas ("Predicted label") para las clases BES, BH y BM.

C. *Validación externa preliminar en campos similares*

Con el fin de evaluar la capacidad de generalización del modelo, se realizó una validación externa preliminar utilizando seis pozos de otros campos con condiciones operativas similares a las del campo Lago Agrio. Para cada pozo, se generó una recomendación del modelo Stacking Classifier y se comparó el sistema de levantamiento reportado en cada caso. La Tabla V presenta la información de los pozos evaluados y los resultados de la comparación de su concordancia. De los pozos evaluados, 5 de 6 concuerdan entre lo instalado y lo predicho por el modelo, lo cual representa una concordancia del 83,3 %, lo que evidencia un desempeño satisfactorio en escenarios fuera del conjunto de entrenamiento. Este resultado se consideró preliminar debido al tamaño limitado de la muestra.

TABLA V
VALIDACIÓN PRELIMINAR EN POZOS: PREDICCIÓN VS. SISTEMA INSTALADO

Pozo	Instalación	Predicción	Concordancia
GTA-02	BH	BH	SI
GTA-15	BH	BH	SI
PRH 1	BH	BH	SI
PRH 43H	BES	BH	NO
PRH 40H	BES	BES	SI
PRH 73H	BES	BES	SI

La única discrepancia se presentó en el pozo PRH 43H donde el documento sostenía un sistema BES con alternativa de cambio en el tipo de bomba, mientras que el modelo predictivo dio como recomendación el uso de BH. Esta diferencia se explicó porque el modelo basó su recomendación únicamente en los patrones estadísticos contenidos en las variables de entrada, mientras que el otro análisis incorporó criterios adicionales de selección orientados a la factibilidad del cambio y el desempeño operativo. Por ello, la no coincidencia en PRH 43H no necesariamente indicó un error del modelo, sino una diferencia entre una recomendación basada en similitud histórica y una decisión de cambio basada en restricciones y objetivos específicos de estudio.

D. *Aplicativo Web*

El aplicativo web OptiLift se utilizó como interfaz de demostración para ejecutar el modelo seleccionado en un entorno operativo, permitiendo transformar la predicción de un resultado consultable por el usuario. El acceso al aplicativo se encuentra disponible en OptiLift. El aplicativo es accesible tanto desde computadoras de escritorio como desde dispositivos móviles mediante un navegador. Su diseño responsivo garantiza la compatibilidad entre plataformas, permitiendo al usuario interactuar con la interfaz y generar predicciones sin requerir instalación local.

La Fig. 7 presenta la pantalla principal del aplicativo, donde se integra el ingreso manual de variables, mientras que

la Fig. 8 muestra el panel de salida con la recomendación generada

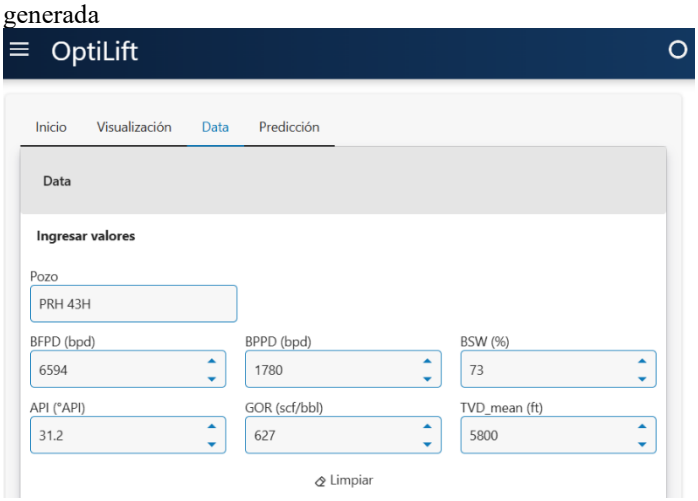

Fig. 7 Pantalla de ingreso de variables del aplicativo web

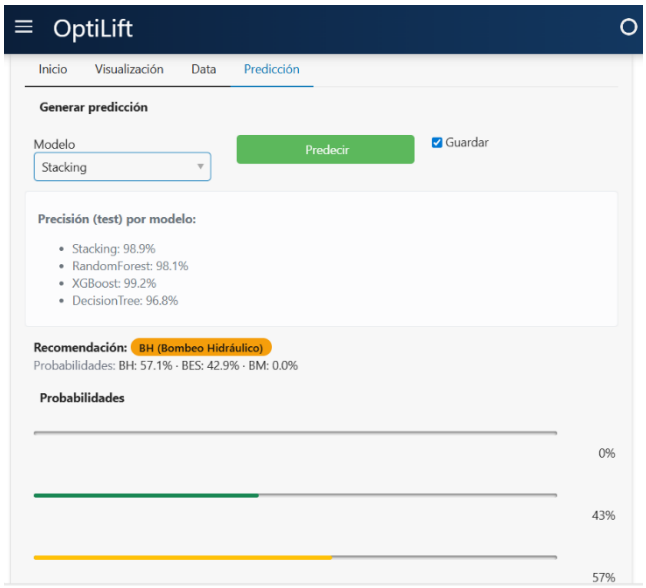

Fig. 8 Pantalla de predicción y recomendación

Durante las pruebas realizadas, el aplicativo permitió procesar registros de pozos de forma directa, generando como salida la clase recomendada de SLA y sus probabilidades estimadas por alternativa. Esto facilitó dos resultados relevantes: identificar el sistema más probable y detectar casos con mayor incertidumbre, cuando las probabilidades entre alternativas son cercanas, los cuales se consideraron para revisión técnica adicional [18].

IV. Conclusiones

En este trabajo se presentó un enfoque basado en aprendizaje automático para apoyar la selección de sistemas de levantamiento artificial a partir de variables operativas y productivas. En la evaluación comparativa, los modelos de ensamble alcanzaron el mejor desempeño en general, con diferencias marginales entre los dos enfoques líderes. Aunque XGBoost mostró una ventaja ligera en el conjunto de prueba, Stacking Classifier exhibió un desempeño competitivo y un comportamiento comparable en validación cruzada, sin evidenciar una brecha train-validación mayor. En consecuencia, se seleccionó Stacking como modelo final debido a su naturaleza de ensamble y su estructura modular, que permite incorporar modelos base adicionales y actualizar el sistema conforme se disponga de mayor volumen y diversidad de datos.

Adicionalmente, la evaluación en pozos de campos con características similares al campo de estudio permitió obtener una concordancia preliminar del 83.3 %, lo que sugirió un desempeño consistente fuera del conjunto de entrenamiento. No obstante, este resultado se interpretó como preliminar debido al tamaño limitado de la muestra.

Como trabajo futuro, se consideró ampliar el número de pozos evaluados e incorporar variables adicionales relacionadas con restricciones operativas y condiciones mecánicas, con el fin de fortalecer la capacidad de generalización y aplicabilidad del modelo como herramienta de apoyo en la toma de decisiones.

Referencias

[1] B. J. J ARPs and M. Aime, ‘Chapter II. Petroleum Economics Analysis of Decline Curves’, May 1944.

[2] Z. Khatib, P. Verbeek, and S. I. E&p, ‘HSE HORIZONS Water to Value-Produced Water Management for Sustainable Field Development of Mature and Green Fields’, 2003.

[3] Garcia C. Andriuska D and Quintero L. Evelyn, ‘Sistema de levantamiento’, 2005.

[4] M. Janadeleh, R. Ghamarpoor, N. Kadhim Abbood, S. Hosseini, H. N. Al-Saedi, and A. Z. Hezave, ‘Evaluation and selection of the best artificial lift method for optimal production using PIPESIM software’, *Heliyon*, vol. 10, no. 17, Sep. 2024, doi: 10.1016/j.heliyon.2024.e36934.

[5] B. Apolo *et al.*, ‘Methodology for the selection of Artificial Survey Systems in oil fields of Ecuador’, Jul. 2020, doi: 10.18687/LACCEI2020.1.1.66.

[6] G. V. Á. Rolando, L. D. J. Rodrigo, and L. B. A. Valentino, ‘Case study: Strategy for selection and optimization of the artificial lift system in eastern Ecuador’, in *Proceedings of the LACCEI international Multi-conference for Engineering, Education and Technology*, Latin American and Caribbean Consortium of Engineering Institutions, 2020. doi: 10.18687/LACCEI2020.1.1.272.

[7] Z. Tariq *et al.*, ‘A systematic review of data science and machine learning applications to the oil and gas industry’, vol. 11, pp. 4339–4374, 2021, doi: 10.1007/s13202-021-01302-2.

[8] M. Alhaj, A. Mahdi, ; M Amish, and G. Oluyemi, ‘Artificial Lift Selection Methods in Conventional and Unconventional Wells: A Summary and Review from Old Techniques to Machine Learning Applications’, vol. 9, no. 3, 2024, doi: 10.38124/ijisrt/IJISRT24MAR2108.

[9] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist, and H. Oza, ‘Application of machine learning and artificial intelligence in oil and gas industry’, Dec. 01, 2021, *KeAi Publishing Communications Ltd.* doi: 10.1016/j.ptlrs.2021.05.009.

[10] F. Pedregosa *et al.*, ‘Scikit-learn: Machine Learning in Python’, *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830,

2011, Accessed: Jul. 05, 2025. [Online]. Available: http://scikit-learn.sourceforge.net.
[11] F. Carrion-Maldonado and J. L. Dávila, 'Machine Learning using synthetic data in the selection of artificial lift systems for oil wells', 2022, doi: 10.18687/LACCEI2022.1.1.616.
[12] T. Ounsakul, T. Sirirattanachatchawan, W. Pattarachupong, Y. Yokrat, and P. Ekkawong, 'IPTC-19423-MS Artificial Lift Selection Using Machine Learning', 2019.
[13] S. Sharma, S. Chatterjee, and D. Tran, 'entropy Winsorization for Robust Bayesian Neural Networks', 2021, doi: 10.3390/e23111546.
[14] G. Chen, B. Hou, and T. Lei, 'A New Monte Carlo sampling method based on Gaussian Mixture Model for imbalanced data classification', 2023, doi: 10.3934/mbe.2023794.
[15] D. H. Wolpert, 'Stacked Generalization', 1992.
[16] L. Breiman, 'Random Forests', 2001.
[17] M. Sokolova and G. Lapalme, 'A systematic analysis of performance measures for classification tasks', *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/j.ipm.2009.03.002.
[18] KELVIN JORGE BARRERA CRUZ, 'UPSE-TIP-2021-0028 (Caracteristica campo Pucuna)', 2021.