

Machine Learning using synthetic data in the selection of artificial lift systems for oil wells

Freddy Carrion-Maldonado, Master¹, Jorge Lliguizaca Dávila, Master¹

¹Escuela Superior Politécnica del Litoral, Ecuador, fpcarrio@espol.edu.ec, jorollig@espol.edu.ec

Abstract– The problem addressed the selection of hydrocarbon artificial lift methods, proposing techniques that can be more effective, whose results present the lowest possible degree of error, which should be automated as much as possible.

The study proposed as a solution the implementation of computational techniques to solve the challenge of selecting artificial lift methods, and in this way presented a new alternative that provides support to the decision-making process, orienting the final criteria to be based on data.

The methodology was implemented in 4 wells of interest, using python to code 4 supervised machine learning algorithms for classification. These models were trained with synthetic data. The training, evaluation, and execution of the algorithms were carried out through the use of the scikit-learn library, the whole project was developed and can be visualized in a [Python IDE that can be accessed online](#).

The selected algorithm was the decision tree algorithm, and its results were consistent with the quality of the training data. In the analysis, the influence of the distribution of training and test data was verified, which for the algorithm with the best performance exhibits an improvement when using the proportion 80 and 20.

Finally, the comparisons made with the in-situ engineering evaluations show that the proposal provides a good data-driven guide, which could even be improved by adding new criteria and an improved database.

Keywords-- Machine Learning, artificial lift selection, classification algorithm, production optimization, data science.

Digital Object Identifier (DOI):

<http://dx.doi.org/10.18687/LACCEI2022.1.1.616>

ISBN: 978-628-95207-0-5 ISSN: 2414-6390

Machine Learning utilizando datos sintéticos en la Selección de Sistemas de Levantamiento Artificial para pozos petroleros

Resumen– El problema planteado abordó la selección de métodos de levantamiento artificial de hidrocarburos, proponiendo técnicas que pueden ser más efectivas, cuyos resultados presenten el menor grado de error posible, las cuales deberían ser en lo posible automatizadas.

El estudio propone como solución a este desafío de selección de métodos de levantamiento artificial mediante la implementación de técnicas computacionales, y de esta forma proponer una nueva alternativa que brinde soporte a la toma de decisiones orientando a que los criterios finales sean basados en datos.

Se implementó la metodología en 4 pozos de interés, utilizando Python para programar 4 algoritmos de machine learning de clasificación. Los cuales, fueron entrenados con datos sintéticos. El entrenamiento, evaluación y ejecución de los algoritmos se lo llevó a cabo mediante el uso de la librería scikit-learn, todo el proyecto fue desarrollado y puede ser visualizado en un IDE de Python al que se puede acceder en línea.

El algoritmo seleccionado fue el de árbol de decisiones y sus resultados fueron consistentes con la calidad de los datos del entrenamiento. En el análisis se verificó la influencia de la distribución de datos de entrenamiento y prueba, que para el algoritmo con el mejor rendimiento exhibe una mejora al usar la proporción 80 y 20.

Finalmente, las comparaciones realizadas con las evaluaciones ingenieriles in situ demuestran que la propuesta proporciona una buena guía basada en los datos, la cual incluso podría ser mejorada añadiendo nuevos criterios y una base de datos mejorada.

Palabras clave — Aprendizaje automático, selección de levantamiento artificial, algoritmos de clasificación, optimización de producción, ciencia de datos.

I. INTRODUCCIÓN

La optimización de la producción involucra la consideración de diferentes aspectos como la supervisión y el control de producción del pozo, la gestión de yacimientos, la optimización del mecanismo de producción artificial, el mantenimiento predictivo del levantamiento artificial, entre otros aspectos que responden a la operatividad de las herramientas, a las características de los reservorios, al entorno económico, entre otros.[1, 2, 3].

En la Cuenca Oriente del Ecuador se han llevado a cabo distintas propuestas para realizar una selección del sistema de levantamiento artificial adecuado para las características de la

cuenca y las necesidades de sus operadoras. Entre ellas se han desarrollado: metodologías de análisis en base a los historiales de producción, esquemas mecánicos y condiciones de yacimiento [4], tablas de doble entrada que establecen parámetros y rangos de operación de los sistemas a utilizarse [5] y una matriz de preselección que combina el índice de productividad y las condiciones mecánicas del pozo[6].

Si bien las metodologías tradicionales tienen validez de aplicación, se pueden aplicar alternativas en una industria de petróleo y gas que enfrenta múltiples desafíos, sobre todo con respecto a la manipulación y procesamiento de grandes cantidades de información en sus procesos de toma de decisiones.

Extensas cantidades de archivos son generados a diario dentro de cada área de la industria petrolera. Por tanto, la digitalización de los flujos de trabajo mediante el “Machine Learning” –aprendizaje automático– y la analítica avanzada es la nueva estrategia para añadir valor empresarial en el sector del petróleo y el gas.

El Machine Learning abiertamente puede ser clasificado en dos fases como la experimentación, que se refiere a los esfuerzos centrados en la preparación de datos, selección de algoritmos y la validación/verificación del modelo. Por otro lado, la parte operacional se refiere al proceso de ejecución del modelo, su monitoreo, consumo y servicios eficientes y medibles [7].

Su aplicación se enfoca en ofrecer soluciones en términos de automatización o mejoramiento de procesos. La existencia de esta ciencia y dominio informático respectivamente, puede eliminar significativamente el factor de riesgo y costos de procesos prácticamente dentro de cualquier área de la industria [8] [9].

En cuanto al número de estudios de aplicación de machine learning estos han ido en crecimiento [7], en cuanto al subsuelo se encuentran estudios orientados a encontrar valores no registrados en los perfiles del pozo [10], en la predicción de parámetros del reservorio [11] e incluso localmente se ha empleado el algoritmo de los K vecinos más cercanos para determinar los métodos de EOR a implementarse [12].

Sin embargo, una pequeña cantidad de datos, la obtención de datos erróneos a causa de la mala interpretación del campo, entre otros factores que afectan a los datos de entrada, puede derivar en una mala calidad de resultados por parte de los algoritmos, y esto, podría convertirse en la barrera más grande que puede presentar la implementación de la inteligencia artificial en una empresa [13] [9].

Machine Learning puede aplicarse a distintas etapas de la producción de pozos petroleros, permitiendo minimizar el costo del ciclo de vida de los pozos de bombeo artificial. En este ámbito, la capacidad de aprendizaje de los algoritmos permite que nuevos datos se incorporen al conjunto de datos existentes y que el modelo responda al cambio dinámico de las condiciones de los yacimientos. [14], brindando soporte sobre el terreno a comprender problemas complejos como la selección del sistema de levantamiento artificial óptimo.

Otros casos de estudio han mostrado que los algoritmos ayudan a los ingenieros a seleccionar el momento más adecuado, para la ejecución de un mantenimiento y optimización en los sistemas de producción [15] y que también podría reducirse una pérdida de producción considerable debido a la precisión en la detección de problemas antes de que se produzca un fallo o daño no programado [16].

Este estudio, cuya metodología se presenta a continuación, tiene como objetivo demostrar que el desafío de selección de métodos de levantamiento artificial puede ser superado mediante la implementación de técnicas computacionales, y de esta forma proponer una nueva alternativa que brinde soporte a la toma de decisiones, la cual, en este punto en particular, se enfoca en tomar decisiones basadas en datos, cuyo uso hoy en día está en tendencia, con el auge de áreas como la ciencia de datos y el machine learning.

II. METODOLOGÍA

La selección del método de levantamiento artificial más adecuado, para implementarlo en algún pozo de interés fue realizada, utilizando algoritmos de machine learning de aprendizaje supervisado de clasificación. Los cuales, fueron entrenados con datos sintéticos.

Haciendo énfasis en la información de parámetros generados, los mismos fueron creados como valores aleatorios, con distribución uniforme, tomando en cuenta como rango, los valores mínimos y máximos de cada parámetro analizado tomando como referencia los valores en la TABLA I.

TABLA I
PARÁMETROS CARACTERÍSTICOS DE LOS SISTEMAS DE LEVANTAMIENTO ARTIFICIAL

	Bombeo Mecánico	Bombeo “Gas Lift”	Bombeo Hidráulico	Bombeo Electro sumergible	Bombeo De Cavidades Progresivas
Parámetros Típicos	Mínimo Promedio Máximo	Mínimo Promedio Máximo	Mínimo Promedio Máximo	Mínimo Promedio Máximo	Mínimo Promedio Máximo
Tasa de producción deseada (bpd)	5 752.5 1500	100 5050 10000	50 2025 4000	100 15050 30000	5 1102.5 2200
Profundidad de levantamiento TVD (pies)	100 8050 16000	5000 10000 15000	5000 10000 15000	5000 10000 15000	2000 6000 10000
Gravedad (API)	>8	>15	>8	>10	<35
Temperatura de yacimiento (°F)	100 225 350	100 250 400	100 300 500	100 250 400	75 162.5 250
Cambio angular limita la Prof. De levantamiento (°)	0 10 20	0 40 80	0 30 60	0 40 80	0 5 10
Grado de inclinación del pozo (°)	0 10 20	0 35 70	0 45 90	0 40 80	0 45 90

Fuente: [17]

Los parámetros utilizados como posibles variables independientes del modelo o también conocidas como datos de entrada, son variables numéricas, típicamente utilizadas para realizar selección de métodos de levantamiento artificial de la manera tradicional, la cual está basada en criterios ingenieriles.

Para este estudio fueron empleados: el caudal de petróleo esperado, la profundidad vertical verdadera del pozo comúnmente conocida como tvd, true vertical depth por sus siglas en inglés, la gravedad API del crudo y la temperatura promedio del yacimiento.

No obstante, en inicio también se consideró como variable de interés, la inclinación del pozo, pero debido a que estos valores se encontraban disponibles, se optó por solamente considerar los 4 parámetros previamente mencionados.

El entrenamiento, evaluación y ejecución de los algoritmos se lo llevó a cabo mediante la aplicación de la librería scikit-learn del lenguaje Python. Entre los que se encuentran aquellos utilizados en este estudio, los cuales fueron: el algoritmo de los k vecinos más cercanos (KNN), el algoritmo de árbol de decisión, el algoritmo clasificador de vectores de soporte, conocido en inglés como support vector classifier (SVC) y el algoritmo de bosques aleatorios, conocido en inglés como randomforest.

Por otro lado, es pertinente mencionar que se escogió algoritmos de aprendizaje supervisado de clasificación porque la variable predictora o variable de salida de estos modelos, es categórica, es decir, son los cinco métodos de levantamiento artificial analizados, los cuales son: el bombeo mecánico, bombeo hidráulico, levantamiento asistido por gas (gas lift, bombeo electro sumergible (BES) y bombeo por cavidades progresivas (PCP).

A. Caso de estudio

Para este estudio, se tomó como referencia a cuatro pozos productores de petróleo, localizados en la región amazónica del Ecuador, los cuales requerían de la implementación de un sistema de levantamiento artificial, para continuar con la producción de hidrocarburos, debido a que su presión de reservorio ya no era lo suficientemente alta, y que

consecuentemente estos puedan continuar produciendo hidrocarburos de forma natural hasta superficie.

TABLA II
PARÁMETROS DE YACIMIENTO DE LOS POZOS ANALIZADOS

Parámetros	Pozo ESPOL 1	Pozo ESPOL 2	Pozo ESPOL 3	Pozo ESPOL 4
Caudal (bpd)	720	166	1071	411.4
Profundidad (pies)	10355	10085	9681	9612
Gravedad API	29.6	30	24.1	30.8
Temperatura (°F)	214	232	222	215

Los parámetros de estos pozos fueron detallados en la tabla I, en donde se encuentran: caudal de petróleo esperado, profundidad vertical verdadera, gravedad API y temperatura. Para efectos prácticos, se denominó a estos pozos como: ESPOL 1, ESPOL 2, ESPOL 3 y ESPOL 4.

B. Definición del problema

El principal desafío que conlleva la selección de métodos de levantamiento artificial de hidrocarburos es la implementación de técnicas más efectivas que generen resultados con el menor grado de error posible, las cuales deberían ser en lo posible automatizadas mediante la inclusión de programas computacionales, debido a esto, en este artículo, se realizó la implementación de cuatro algoritmos de machine-learning, para la realización de dicha selección.

C. Procedimiento

El procedimiento siguió los pasos del flujo de trabajo sugerido para todo proyecto de machine learning, el cual se lo suele conocer como machine learning pipeline como se muestra en la figura 1.

De la figura 1 se puede apreciar el flujo de trabajo resumido de la metodología implementada en este estudio. Cabe recalcar, que el paso 8 a pesar de que no fue nombrado en este procedimiento, es muy necesario de implementarlo, con la finalidad de monitorear el rendimiento del modelo.

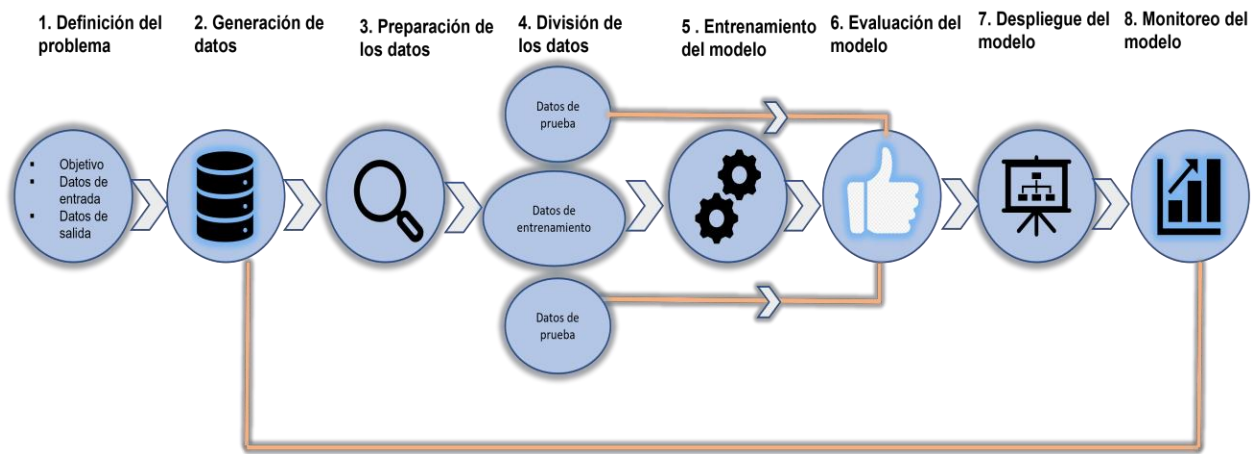


Figura 1. Flujo de trabajo de un proyecto de machine learning (machine learning pipeline)

Paso 1. Definir el problema e identificar que tipos de algoritmos de machine learning se pueden utilizar.

Paso 2. Crear los datos aleatorios sintéticos y transformarlos en datos tabulares (dataframe de Python), utilizando los rangos permitidos (mínimo y máximo) de las propiedades de yacimiento consideras para este estudio (ver tabla I). En total se generaron 1139 observaciones, distribuyéndose de la siguiente forma: 250 pertenecen a bombeo mecánico, 200 pertenecen al levantamiento asistido por gas (gas lift), 165 pertenecen al bombeo hidráulico, 325 pertenecen al bombeo electro sumergible (ESP) y 200 pertenecen al bombeo por cavidades progresivas (PCP).

Paso 3. Realizar el análisis exploratorio de los datos generados, con la finalidad de identificar posibles datos nulos, revisar las correlaciones entre las variables, visualizar las distribuciones estadísticas de las variables mediante gráficos de dispersión entre variables como se presenta en la figura 2; y posteriormente realizar el procesamiento requerido de esta información, con la finalidad de que quede lista para ingresarla a los modelos predictivos.

Paso 4. Separar el conjunto de datos en conjuntos de prueba y de entrenamiento, en donde para este proyecto, se ha separado los datos en 3 distribuciones distintas y así evaluar la sensibilidad a este paso, las cuales fueron: 70% para entrenamiento y 30% para prueba correspondiente a la distribución 1, 75% para entrenamiento y 25% para prueba correspondientes a la distribución 2 y 80% para entrenamiento y 20% para prueba, con la finalidad de identificar la distribución que permite un mejor entrenamiento de los modelos. Esta separación de los datos fue implementada utilizando la librería de Python scikit-learn.

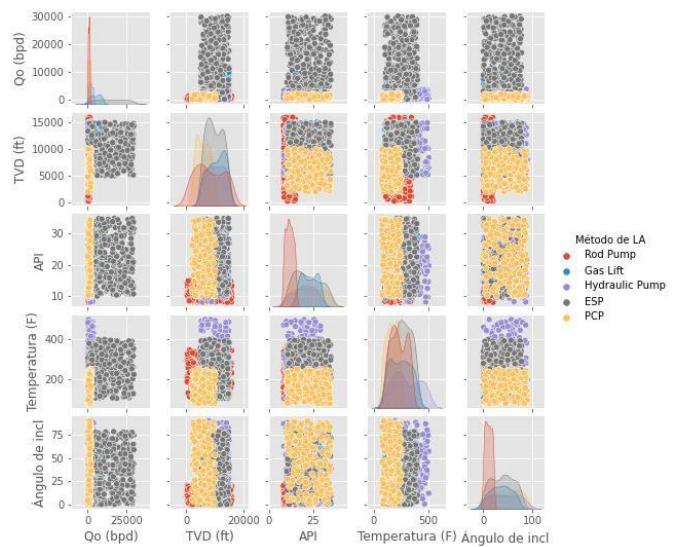


Figura 2. Gráficos de dispersión entre variables

Paso 5. Posteriormente, se procede a entrenar los modelos, utilizando solamente los datos de entrenamiento, así como los hiperparámetros más importantes de cada algoritmo, los cuáles son los argumentos con los que funcionan los algoritmos, cuyos valores óptimos fueron escogidos utilizando un procedimiento de ajuste, el cual se enfoca en identificar sus valores óptimos que permiten un mejor entrenamiento de los algoritmos analizados.

Paso 6. Una vez que los algoritmos han sido entrenados, se procede a evaluar su rendimiento, utilizando el conjunto de datos que los algoritmos no han visto previamente, es decir, para esta evaluación, se utilizaron los datos de prueba.

Paso 7. Finalmente, se compararon los rendimientos del entrenamiento de cada algoritmo, con el objetivo de seleccionar el algoritmo que tengo una mejor generalización para futuros datos de entrada. Es decir, que se pueda utilizar el

modelo para realizar la selección del método de levantamiento artificial de petróleo para los pozos que se requiera analizar.

III. ANÁLISIS Y RESULTADOS

En esta sección, primero se realizó una verificación de los parámetros utilizados como variables de entrada de los modelos, con la finalidad de comprobar sus correlaciones e identificar posibles casos de multicolinealidad.

Por consiguiente, se construyó una matriz de correlaciones (correlación de Pearson) que presenta sus resultados en la figura 3, en esta se observa que las correlaciones entre estos parámetros son en general bajas, lo cual es beneficioso para nuestros modelos, debido a que, con esto, se descarta la presencia de multicolinealidad, la cual podría ocurrir si estos valores se encontrasen en el rango de 70 a 80%. Por lo tanto, se puede concluir que las variables seleccionadas son adecuadas.

Es importante resaltar, que a pesar de que los resultados de correlación entre la variable ángulo de inclinación con el resto de los parámetros es también aceptable, no se la incluyó en el entrenamiento de los modelos, debido a la frecuente falta de información sobre este dato.

Por otro lado, para entrenar los datos generados, se utilizaron algoritmos de machine learning de aprendizaje supervisado de clasificación, debido a que la variable predictora (métodos de levantamiento artificial) es categórica. Además, es pertinente mencionar que, los modelos fueron evaluados mediante el cálculo de la métrica precisión, conocida como accuracy en inglés, la cual es utilizada para algoritmos cuyas variables predictoras son categóricas.

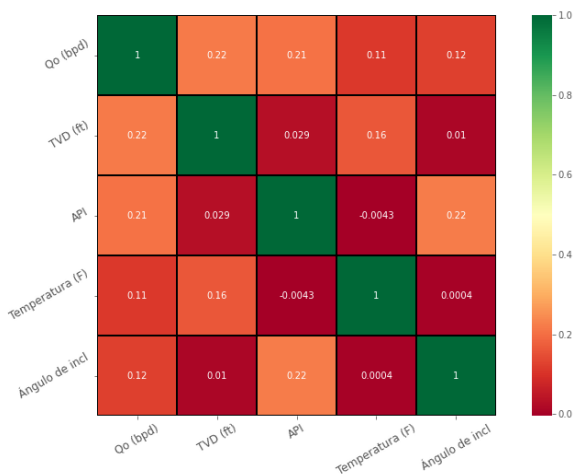


Figura 3. Matriz de correlación

A. Comparación de los rendimientos de los algoritmos

A partir de la TABLA III, se visualiza el resumen de los resultados del entrenamiento de los cuatro algoritmos utilizados para este estudio, considerando tres distribuciones distintas para la distribución de los datos de entrenamiento y prueba.

TABLA III
LOS RENDIMIENTOS DE LOS ALGORITMOS POR DISTRIBUCIÓN

Algoritmo	Accuracy distribución 1	Accuracy distribución 2	Accuracy distribución 3
KNN	0.65	0.64	0.63
Decision tree	0.78	0.78	0.80
SVC	0.68	0.65	0.64
Random Forest	0.58	0.58	0.59

En base a esta información, se dedujo que el modelo que presenta el mejor desempeño es el implementado con el algoritmo de árbol de decisión, utilizando la distribución 3 (80% para entrenamiento y 20% para prueba), debido a que obtuvo la métrica de precisión más alta (aproximadamente 80%).

Se observó que las distribuciones de los datos de prueba y entrenamiento influyen en el mejoramiento del rendimiento de algoritmos como el decision tree y random forest, mientras que en el caso de los KNN y SVC sus rendimientos se ven disminuidos.

Finalmente, los resultados obtenidos demuestran que el algoritmo ha sido correctamente entrenado y es apto para realizar la predicción del método de levantamiento artificial de hidrocarburos recomendado para cada uno de los pozos analizados.

B. Análisis del desempeño del algoritmo de los K vecinos más cercanos (KNN por sus siglas en inglés)

La evaluación para este modelo fue realizada utilizando varios valores de K, ya que este es el parámetro más influyente del algoritmo. En total se tomaron en cuenta 30 valores de K, del 1 al 30.

De acuerdo con los resultados visualizados a partir de la figura 4, donde fueron graficadas curvas de la métrica accuracy vs el número de k vecinos, utilizando los datos de prueba y entrenamiento, con el fin de determinar el desempeño del algoritmo para cada conjunto de datos, se puede identificar que cuando el valor de K es 9, es decir, cuando se toman 9 valores alrededor del valor a evaluar, el algoritmo presenta su mejor rendimiento, debido a que a estas condiciones su precisión de la curva de prueba es de aproximadamente 64%.

Este número de k vecinos fue escogido como el más adecuado, debido a que, con valores mayores a este, la curva de prueba cambia su tendencia, lo cual no es beneficioso para medir el desempeño del modelo.

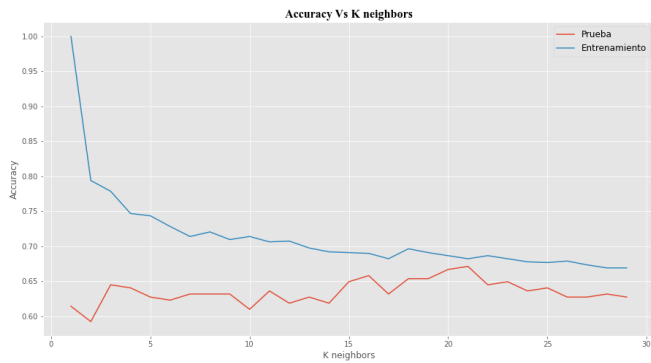


Figura 4. Comportamiento del algoritmo KNN

C. Análisis del desempeño del algoritmo de árbol de decisión (Decision Tree en inglés)

La evaluación para este modelo fue realizada utilizando varios valores de max depth (máxima amplitud en español), ya que este es uno de los parámetros más influyente del algoritmo. En total se tomaron en cuenta 18 valores de max depth, del 2 al 19.

De acuerdo con los resultados visualizados a partir de la figura 5, se puede identificar que cuando el valor de max depth es 5, el algoritmo presenta su mejor rendimiento, debido a que, a estas condiciones, la precisión de la curva de prueba es de aproximadamente 80%.

Este valor de max depth fue escogido como el más adecuado, debido a que, con valores mayores a este, la curva de prueba cambia su tendencia, lo cual no es beneficioso para medir el desempeño del modelo, ya que de esta forma el modelo estaría sobre ajustado, lo cual afecta su rendimiento.

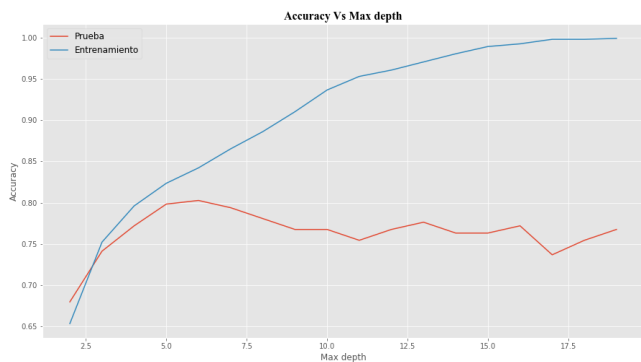


Figura 5. Comportamiento del algoritmo árbol de decisión

D. Análisis del desempeño del algoritmo clasificador de vectores de soporte (Support Vector Classifier (SVC) en inglés)

La evaluación para este modelo fue realizada utilizando varios valores de C, el es un parámetro que actúa como penalizador, evitando la clasificación errónea de cada ejemplo de entrenamiento. Además, este es uno de los parámetros más influyentes del algoritmo. Se tomaron en cuenta los siguientes valores para el entrenamiento de este modelo: 0.001, 0.01, 0.1, 1, 10, 100

De acuerdo con los resultados visualizados a partir de la figura 6, se puede identificar que cuando el valor de C es 100, el algoritmo presenta su mejor rendimiento, debido a que a estas condiciones su precisión de la curva de prueba es de aproximadamente 64%.

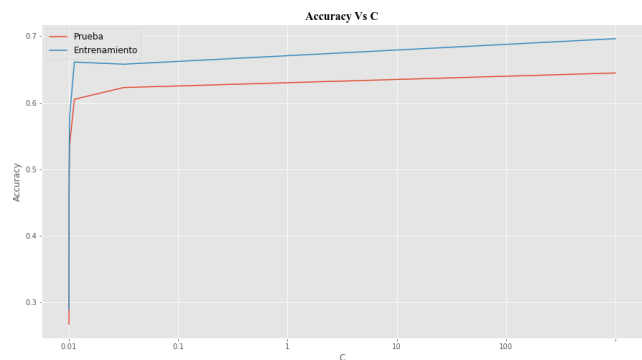


Figura 6. Comportamiento del algoritmo clasificador de vectores de soporte (SVC por sus siglas en inglés)

E. Análisis del desempeño del algoritmo bosque aleatorio (Random Forest) en inglés)

La evaluación para este modelo fue realizada utilizando varias cantidades para el max depth (máxima amplitud en español), ya que este es uno de los parámetros más influyentes del algoritmo. En total se tomaron en cuenta 10 valores de max depth, desde el 0.1 al 1, incrementando los valores en escala de 0.1.

De acuerdo con los resultados visualizados a partir de la figura 7, se puede identificar que cuando el valor de max depth es 1, el algoritmo presenta su mejor rendimiento, debido a que a estas condiciones su precisión de la curva de prueba es de aproximadamente 58%.

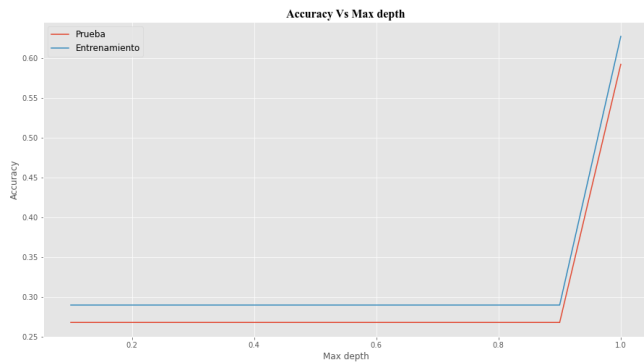


Figura 7. Comportamiento del algoritmo bosque aleatorio (RandomForest en inglés)

Es importante resaltar que, los resultados mostrados en las figuras 4, 5, 6 y 7 fueron obtenidos, tomando en cuenta la distribución 2 de los datos, con la cual, la distribución de estos es de 80% para entrenamiento y 20% para prueba.

F. Comparación de las selecciones realizadas por el algoritmo por cada distribución de datos

TABLA IV
SELECCIÓN DE SISTEMAS DE LEVANTAMIENTO ARTIFICIAL POR DISTRIBUCIÓN

Pozo	Resultados distribución 1	Resultados distribución 2	Resultados distribución 3	Resultados evaluación tradicional
ESPOL 1	BH	BH	BH	BH
ESPOL 2	BH	BH	BH	BES*
ESPOL 3	BCP	BH	BH	BES
ESPOL 4	BCP	BES	BES	BES

Los resultados mostrados en la Tabla IV muestran diferentes opciones tales como el Bombeo Hidráulico BH, el Bombeo de Cavidades Progresivas BCP y el Bombeo Electrosumergible para cada uno de los pozos correspondientes al uso de cada distribución de datos de entrenamiento y prueba.

Para el pozo ESPOL 1 los resultados coinciden para las tres distribuciones e incluso es confirmado el resultado con la evaluación realizada en campo mediante la evaluación tradicional.

El pozo ESPOL 2 presenta coincidencias en cuanto las distribuciones sin embargo el resultado final no es similar al realizado por la evaluación de campo, esto se podría justificar porque en este caso la decisión final determinó que si bien el bombeo hidráulico era la mejor opción bajo análisis técnico, al incluir los factores económicos finalmente se implementó el sistema de bombeo electrosumergible.

Nuestro tercer caso de análisis presentó un cambio en cuanto a la decisión del algoritmo, anteriormente se determinó

que se mejora su decisión al aumentar la proporción de datos de entrenamiento, con esto podemos decir que el cambio en su resultado responde a esta mejora.

Al comparar los resultados de ESPOL 3 con la decisión final en campo, se determinó que el método BES era mejor al BH que determinaba el algoritmo, la decisión tomada en campo presenta una congruencia por el gran potencial que presentaba este pozo.

Finalmente el pozo 4 si bien una de sus decisiones es distinta, lo cual es congruente con la relación entre el rendimiento del algoritmo y la distribución, comparando el resultado de la tercera distribución este coincide con la evaluación in situ.

IV. CONCLUSIONES

Los métodos que involucran al Machine Learning proveen una relación entre variables de ingreso para poder predecir con la más alta exactitud un posible resultado. Sin embargo, la exactitud de tal resultado depende enormemente la profundidad del análisis previo y el preprocesamiento de los datos obtenidos en los distintos procesos llevados a cabo; ya que, por lo general, los datos asociados suelen ser extensos y confusos.

Ante lo antes mencionado una primera reflexión o mejora a este trabajo es utilizar datos históricos de campo obtenidos de una gran cantidad de pozos con rendimientos exitosos y no exitosos.

Los resultados han demostrado que al utilizar un amplio conjunto de datos es posible crear un sistema de soporte de decisiones, que permite aportar una recomendación a conocimientos de ingeniería, mediante el entrenamiento de los algoritmos de ML.

Los resultados del pozo 2 y 3 nos permitieron planteamos mejoras para el entrenamiento del algoritmo, aumentando la cantidad de parámetros para el entrenamiento sean estos de tipo técnico o económico.

Además, el conocimiento del dominio también puede utilizarse como mecanismo de retroalimentación para evaluar el rendimiento de un algoritmo y mejorar la precisión de las predicciones.

El algoritmo fue capaz de ofrecer tres resultados que desde un punto de vista técnico fueron acertados corroborando con lo determinado en la locación de producción con el equipo de expertos, y así mismo uno de ellos no coincidió con lo que se determinó en campo.

REFERENCIAS

- [1] J. F. Lea and H. V. Nickens, "Selection of artificial lift," *Soc. Pet. Eng. - SPE Mid-Continent Oper. Symp. 1999, MCOS 1999*, 1999, doi: 10.2118/52157-ms.
- [2] J. Valbuena, E. Pereyra, and C. Sarica, "Defining the artificial lift system selection guidelines for horizontal wells," in *Society of Petroleum Engineers - SPE North America Artificial Lift Conference and Exhibition 2016*, 2016, doi: 10.2118/181229-ms.
- [3] A. Darvish Sarvestani and A. Hadipour, "artificial lift method selection for mature oil fields: A case study," in *Society of Petroleum Engineers - SPE Annual Caspian Technical Conference 2019, CTC 2019*, 2019, doi: 10.2118/198424-ms.
- [4] 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 | Caso de estudio: Estrategia de análisis y selección del Sistema De Levantamiento Artificial para la optimización de un pozo petrolero en el oriente ecua," in *Proceedings of the LACCEI international Multi-conference for Engineering, Education and Technology*, 2020, doi: 10.18687/LACCEI2020.1.1.272.
- [5] J. Z. Anchundia, M. S. Saltos, and F. S. Sares, "Selección del Sistema de levantamiento artificial en el Campo Joma- Distrito oriente ecuatoriano," *Proc. LACCEI Int. Multi-conference Eng. Educ. Technol.*, vol. 2018-July, no. July, pp. 19–21, 2018, doi: 10.18687/LACCEI2018.1.1.315.
- [6] B. Apolo *et al.*, "Methodology for the selection of artificial survey systems in oil fields of Ecuador | Metodología para selección de Sistemas de Levantamiento Artificial en campos petroleros del Ecuador Methodology for the selection of Artificial Survey Systems in oil fie," in *Proceedings of the LACCEI international Multi-conference for Engineering, Education and Technology*, 2020, doi: 10.18687/LACCEI2020.1.1.66.
- [7] Y. Hajizadeh, "Machine learning in oil and gas; a SWOT analysis approach," *J. Pet. Sci. Eng.*, vol. 176, pp. 661–663, May 2019, doi: 10.1016/j.petrol.2019.01.113.
- [8] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist, and H. Oza, "Application of machine learning and artificial intelligence in oil and gas industry," *Pet. Res.*, no. xxxx, 2021, doi: 10.1016/j.ptlrs.2021.05.009.
- [9] K. Balaji *et al.*, "Status of data-driven methods and their applications in oil and gas industry," *Soc. Pet. Eng. - SPE Eur. Featur. 80th EAGE Conf. Exhib. 2018*, 2018, doi: 10.2118/190812-ms.
- [10] H. Jian, L. Chenghui, C. Zhimin, and M. Haiwei, "Integration of deep neural networks and ensemble learning machines for missing well logs estimation," *Flow Meas. Instrum.*, vol. 73, p. 101748, Jun. 2020, doi: 10.1016/j.flowmeasinst.2020.101748.
- [11] D. A. Otchere, T. O. Arbi Ganat, R. Gholami, and S. Ridha, "Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models," *Journal of Petroleum Science and Engineering*, vol. 200. Elsevier B.V., p. 108182, 01-May-2021, doi: 10.1016/j.petrol.2020.108182.
- [12] C. M. F. Paúl, L. D. J. Rodrigo, and P. M. D. Leonardo, "Screening enhanced oil recovery methods using the machine-learning algorithm K Nearest Neighbors (KNN) | Selección de métodos de recuperación mejorada de petróleo utilizando el algoritmo de machine-learning de los K vecinos más cercanos (KNN)," in *Proceedings of the LACCEI international Multi-conference for Engineering, Education and Technology*, 2020, doi: 10.18687/LACCEI2020.1.1.564.
- [13] D. Koroteev and Z. Tekic, "Artificial intelligence in oil and gas upstream: Trends, challenges, and scenarios for the future," *Energy AI*, vol. 3, p. 100041, Mar. 2021, doi: 10.1016/j.egyai.2020.100041.
- [14] T. Ounsakul, T. Sirirattanachatchawan, W. Pattarachupong, Y. Yokrat, and P. Ekkawong, "Artificial lift selection using machine learning," *Int. Pet. Technol. Conf. 2019, IPTC 2019*, 2019.
- [15] F. I. Syed, M. Alshamsi, A. K. Dahaghi, and S. Neghabhan, "Artificial lift system optimization using machine learning applications," *Petroleum*, 2020, doi: 10.1016/j.petlm.2020.08.003.
- [16] P. Bangert, "Predicting and detecting equipment malfunctions using machine learning," *OTC Bras. 2017*, no. October, pp. 1079–1082, 2017, doi: 10.4043/28109-ms.
- [17] Weatherford, "Introduction to Artificial Lift.," 2010.