

Smart Solution To Selecting Artificial Lift And Improve Oil Production Using Machine Learning

Soluciones inteligentes para seleccionar el sistema de levantamiento artificial y mejorar la producción de petróleo mediante el aprendizaje automático

Romel Erazo-Bone, MSc^{1,2}, Fidel Chuchuca-Aguilar, MSc.^{1,2}, and Jefferson Chiquito-Mero, Eng¹, and Emily Patricia Guevara Chávez, Eng¹, Roberto Silva-Zea³

¹Universidad Estatal Península de Santa Elena, Ecuador, raerazo@upse.edu.ec, jchiquito@upse.edu.ec

² Red Internacional de Investigación de Actividades Hidrocarburíferas y Energéticas (RIIAHE), Ecuador, fchuchuca@upse.edu.ec, eguevara@upse.edu.ec

³ Hidroingeniería S.A., Guayaquil, Ecuador, Roberto.silva.zea2@gmail.com

Abstract— The main goal of this research is to evaluate the artificial systems used to produce oil through the machine learning technique with the purpose of selecting a single optimal SLA of section 66 of the Gustavo Galindo Velasco oilfield, to increase production, improve operations, reduce the risks of losses and evaluate the viability and feasibility of the proposal. The present work is based on a descriptive, quantitative and comparative investigation of the different methods of artificial lifting with which it is being produced in the GGV oilfield "section 66". Theoretical research is necessary for the collection of information on the SLAs by BM, HL and plunger or SW. The prediction of the decision tree algorithm gives as a result that the optimal artificial lift system to be implemented in the entire section 66 is mechanical pumping, because the wells that operate with this system obtained a higher production rate during the 2016-2020 period in comparison. to the wells that produce with other artificial lift systems, and in turn generated lower operating and maintenance costs because the interventions carried out in the wells are low cost, in addition money is saved in fuel, grease, lubricants and spare parts .

Keywords—Machine learning, algorithm, production, artificial lift, Oil.

Resumen.- Esta investigación tiene como objetivo evaluar los sistemas a través de la técnica machine learning con el propósito de seleccionar un único SLA óptimo de la sección 66 del campo Gustavo Galindo Velasco, para aumentar la producción, mejorar las operaciones, reducir los riesgos de pérdidas y evaluar la viabilidad y factibilidad de la propuesta. El presente trabajo se basa en una investigación descriptiva, cuantitativa y comparativa de los diferentes métodos de levantamiento artificial con los que se está produciendo en el campo GGV "sección 66". La investigación teórica es necesaria para la recopilación de información sobre los SLA por BM, HL y Pistoneo o SW. La predicción del algoritmo árbol de decisión da como resultado que

el sistema de levantamiento artificial óptimo a implementar en toda la sección 66 es bombeo mecánico, porque los pozos que operan con este sistema obtuvieron una mayor tasa de producción durante el periodo 2016-2020 en comparación a los pozos que producen con otros sistemas de levantamiento artificial, y a su vez generaron menor gasto de operación y mantenimiento debido a que las intervenciones que se realizan en los pozos son de bajo costo, además se economiza dinero en combustible, grasas, lubricantes y repuestos.

Palabras claves: Machine learning, algorithm, producción, Levantamiento Artificial, Petróleo

I. INTRODUCTION

En la actualidad la industria del petróleo busca mejoras continuas para maximizar la tasa de producción y minimizar los costos de operación, al iniciar la etapa de explotación de un yacimiento la producción del petróleo se obtiene por flujo natural, sin embargo, a medida que declina la presión del yacimiento la producción disminuye lo que implica suministrar energía externa para levantar el fluido a superficie, para aquello se utiliza en los pozos un sistema de levantamiento artificial que permita seguir produciendo eficientemente, el mismo que se seleccionará después de realizar un estudio de factibilidad.

La industria petrolera emplea métodos tradicionales de control de procesos, los cuales están sujetos a tener grandes márgenes de error humano, además estos procesos demandan mucho tiempo. Sin embargo, con la aplicación de la técnica machine learning (ML) basada en la creación de algoritmos que aprenden iterativamente dentro de grandes conjuntos de datos (conocido como dataset) permiten a una máquina o software identificar patrones de comportamiento entre millones de datos, realizando estimaciones con un alto grado de precisión de manera rápida y eficiente.

Digital Object Identifier (DOI):

<http://dx.doi.org/10.18687/LACCEI2022.1.1.393>

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

Esta investigación busca evaluar los sistemas a través de la técnica ML con el propósito de seleccionar un único SLA óptimo de la sección 66 del campo GGV, para aumentar la producción, mejorar las operaciones, reducir los riesgos de pérdidas y evaluar la viabilidad y factibilidad de los proyectos.

Un algoritmo de ML determina el SLA óptimo para toda la sección 66, con un análisis crítico comparativo de las variables; condiciones de los pozos, historial de producción, historial de Workover, tipos de fluido en función de su gravedad, costos operativos y mantenimiento.

El principal objetivo de esta investigación es realizar un estudio de factibilidad que permita evaluar los sistemas de levantamiento artificial mediante la aplicación de inteligencia artificial, para conocer cuál de los métodos es el óptimo a implementar en un futuro, en la sección 66 del campo GGV.

En la búsqueda de soluciones a los diferentes problemas durante el proceso de producción de petróleo [1], establecieron el diseño de un algoritmo mediante machine learning para optimizar la producción de finos, y seleccionar el sistema de levantamiento artificial con la finalidad de reducir costos. Con la finalidad realizar mejores predicciones de producción de petróleo utilizando ML [2] propone un algoritmo to develop a correlation that can accurately predict oil rate in artificial gas lift wells, to provide a solution that is simple, easy to use and universally applicable. Durante la vida productiva de un pozo los sistemas de levantamiento sufren fallas que son difíciles de determinar, para disminuir la incertidumbre en la determinación de problemas [3], [4] proponen un algoritmo para detectar fallas en los equipos de fondo y optimizar el SLA utilizando ML.

Existen diferentes tipos de SLA utilizado en la industria del petróleo, entre ellos uno de los más utilizados es el bombeo mecánico (BM), el cual tiene muchas partes móviles las cuales fácilmente tienden a presentar fallas [5], [6] propone el uso de ML para clasificar las cartas dinámicas.

El uso de los sistemas de bombeo electro sumergibles son ampliamente utilizados, sin embargo solucionar fallas implica levantar todo el ensamblaje de fondo, por lo que [7] propone el uso de ML para predecir y reducir las fallas en el sistema.

A. Descripción de los sistemas de levantamiento artificial

El **BM** es uno de los SLA más utilizado más antiguos y utilizado en la industria del petróleo. Utiliza una unidad de bombeo con la cual transmite movimiento a la bomba subsuperficial a través de un conjunto de varillas, y la energía es suministrada por un motor.

Pistoneo o Swab es una técnica de extracción de fluidos del interior de un pozo, desde una profundidad dada hacia la

superficie. Esta operación se realiza debido a que la presión de la formación del pozo no es lo suficientemente mayor como para vencer la columna hidrostática que existe y poder expulsar el líquido drenado a la superficie. Para hacer producir estos pozos, se utiliza cable de acero que va enrollado en el winche o tambor de SW de la unidad de servicio de pozos (tubing o casing) y con ayuda de otros componentes y accesorios hacen posible la recuperación de fluidos.

Herramienta local es el método más sencillo y utilizado por su bajo costo de operación y rápida recuperación de producción en el proceso de extracción de crudo en el Campo Ancón es el sistema de HL [8].

Es una técnica de extracción de fluidos del interior de un pozo, desde una profundidad dada hacia la superficie por medio de un barril con válvula (botella) que es introducida en el pozo.

B. Machine learning

El Aprendizaje automático es un subcampo de la IA centrado en un conjunto de técnicas que permite a las máquinas aprender automáticamente, sin ser programadas explícitamente. El Aprendizaje automático está basado en la creación de algoritmos que aprenden iterativamente dentro de grandes conjuntos de datos (conocido como dataset), a través de un entrenamiento, detectando patrones en los datos y luego adquirir experiencia permitiéndole usar los patrones descubiertos para predecir resultados sin la ayuda de seres humanos. Sin embargo, el aprendizaje automático no es un proceso sencillo, ya que requiere específicamente de información histórica recolectada para describir mejor los datos y en base al conocimiento adquirido mejorar el desempeño de las predicciones, minimizando el error [9].

La figura 1 muestra la interrelación que existen entre la Inteligencia artificial, machine learning y Deep learning, para el caso de nuestro estudio nos enfocaremos en el desarrollo de algoritmos mediante ML.

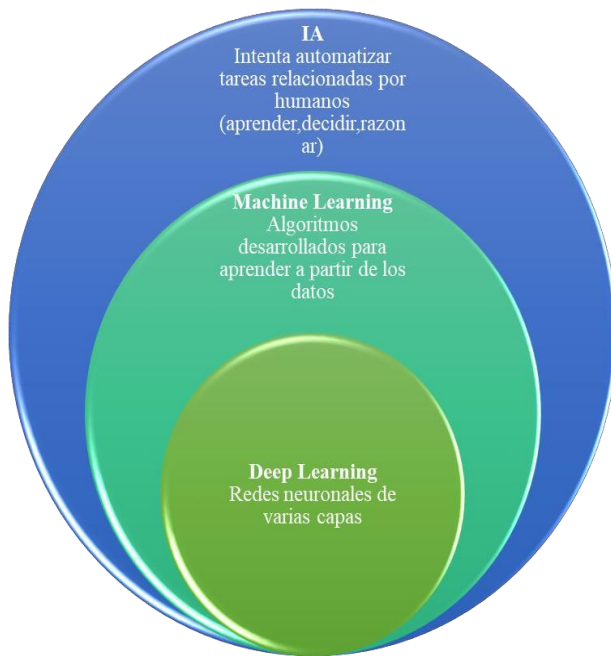


Figura 1. Machine learning subcampo de la IA (Alonso, 2020)

C. Metodología

Para poder realizar la selección del SLA mediante ML, se establecieron las siguientes actividades:

- Un análisis cualitativo de cada uno de los SLA que son utilizados en el área de estudio en función de las diferentes características y generalidades que presentan.
- El uso de un algoritmo de machine learning el cual analizará todos los datos recolectados en el área de estudio: historial de producción, costos operativos y de mantenimiento, gravedad API, ciclos, número de intervenciones.
- Analizar el resultado obtenido del algoritmo con respecto a la selección del SLA óptimo.
- Calcular la factibilidad económica del proyecto en base al análisis de los pozos con mayor producción diaria.

Para la selección del SLA óptimo a implementar en la sección 66 se establecieron los siguientes criterios:

Rango de profundidad. A continuación, se muestran los rangos de profundidad. Los valores de profundidad varían dependiendo de la región en que se encuentre la formación productora, el rango que indican las siguientes tablas son de 2802' – 5209' correspondiente al sistema por BM, el método de HL es de 806' – 4190' y 910' – 5423' para el sistema por Pistoneo.

Tipo de fluido. El petróleo es clasificado en liviano, mediano, pesado y extrapesado de acuerdo con su medición °API, que

diferencia las calidades de crudo, los mismos que están en un rango de 35 a 42°API.

Producción diaria. A continuación, se muestran los datos de la producción diaria promedio de los pozos con su respectivo SLA.

Tabla 1

Producción promedio diaria de los pozos que operan por BM (Pacifpetrol, 2021)	
Nombre del pozo	Producción diaria (bpd)
ANC0437	0,46
ANC0446	1,83
ANC0458	0,92
ANC0461	1,43
ANC0465	4,66
ANC0504	3,85
ANC0508	0,96
ANC0585	3,04
ANC0656	0,96
ANC0778	2,51

A continuación, en la . se detalla la producción diaria en barriles de cada SLA que opera en dicha sección, la formación de la cual se está produciendo y la profundidad del pozo.

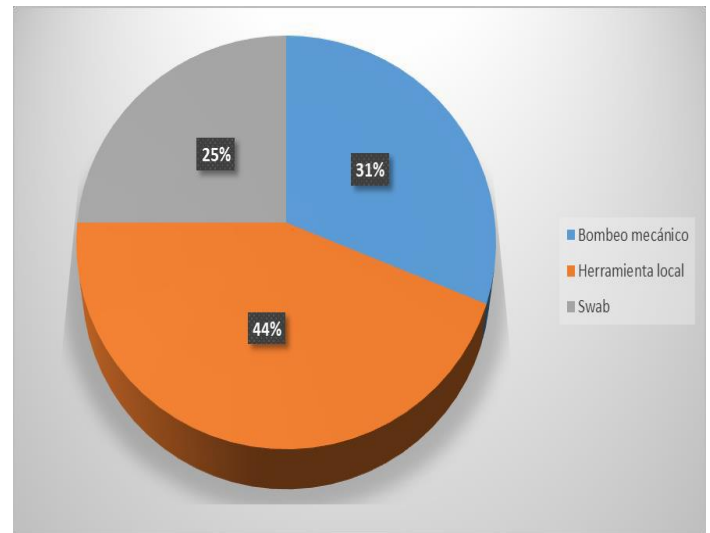


Figura 1. Porcentual de SLA usados en la sección 66 (Pacifpetrol, 2021)

Gastos promedio de los SLA. A continuación, se muestran los gastos promedios mensuales que generan la extracción del crudo y el mantenimiento de los SLA por cada pozo [10].

Tabla 2.
Gasto promedio mensual del SLA BM (por cada pozo)
(Pacifpetrol, 2021)

Ítems	Descripción	Monto \$
Extracción del crudo	Combustible	16,11
	Grasas y lubricantes	25,69
Mantenimiento de la unidad	Mantenimiento de equipos	60,26
TOTAL		102,06

D. Proceso de diseño del algoritmo

El flujo de trabajo establecido por ML comprende las siguientes etapas tal como se muestra en la figura 2:

- **Dataset:** El elemento principal y de mayor importancia es el conjunto de datos y si no se cuenta con una gran cantidad de muestras, difícilmente se podrá realizar el estudio o proyecto.
- **Visualización de Datos:** Generalmente se depuran los datos que no se relacionan a los objetivos propuestos para el estudio o proyecto.
- **Preprocesamiento de los datos:** Se analizan los datos de entrenamiento y se deben dominar conceptualmente los algoritmos estudiados en el Machine Learning y detectar que tipos de entrada de datos reciben, que resultados generan y que tipo de tareas resuelven.
- **Entrenamiento de los datos:** Seleccionado el algoritmo analizado en base a su velocidad de entrenamiento, uso de memoria, precisión predictiva y transparencia, las máquinas ajustarán los parámetros del modelo matemático, para poder en el futuro tomar decisiones por sí mismas. El formato del algoritmo debe contener variables de entrada y de salida, colocar los datos en formas matriciales, crear categorías binarias, y partir y aleatorizar las muestras.

- **Ajustar el modelo:** La función del ajuste dependerá del algoritmo elegido, se normalizan los datos de prueba
- **Prueba del algoritmo:** Evaluar correctamente el modelo y asegurar que funcionará con datos nuevos. Se considera la fase más larga y donde se juega con el ensayo y error [11].
- **Interpretación y evaluación de los Resultados.**

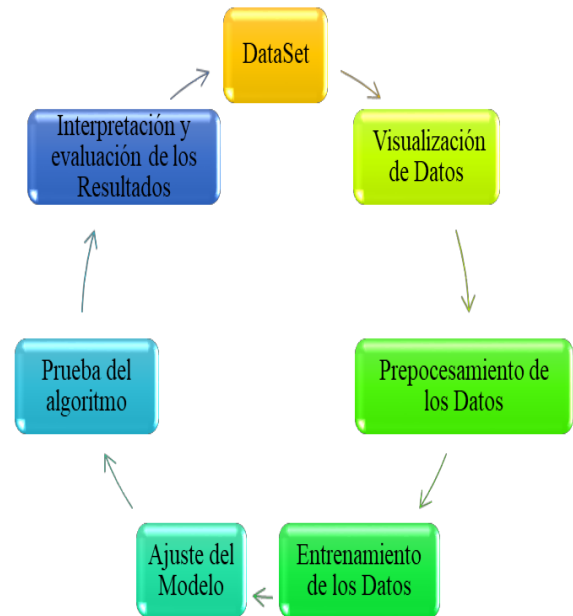


Figura 2. Flujo de trabajo de Machine Learning

De acuerdo a las diferentes técnicas que utiliza ML, para el desarrollo de este proyecto se utilizara el algoritmo de solución de decisión, el cuáles un modelo predictivo que divide el espacio de los predictores agrupando observaciones con valores similares para representar los posibles resultados que ocasionaría la toma de una decisión. Convirtiéndose en una herramienta de apoyo en eventualidades que se presenten, como los costos de recursos y la utilidad.

El algoritmo de árbol de decisión genera un modelo de minería de datos mediante la creación de una serie de divisiones similares a un diagrama de flujo. Estas divisiones se representan como nodos. El algoritmo agrega un nodo al modelo cada vez que una columna de entrada tiene una correlación significativa con la columna de predicción. La forma en que el algoritmo determina una división varía en función de si predice una columna continua o una columna discreta. Se utiliza la selección de características para guiar la selección de los atributos más útiles, donde cada nodo

representa una prueba en una variable específica y las ramas representan los resultados [12]. Ver Figura 3.

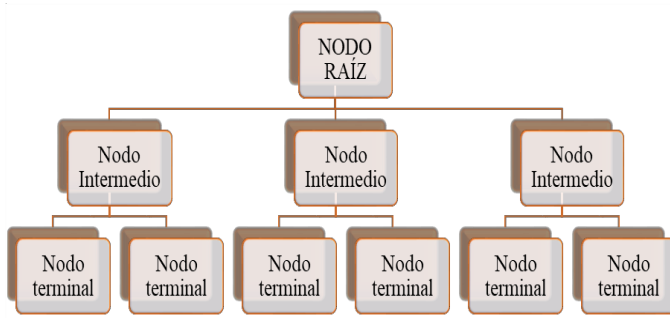


Figura 3. Los árboles de decisión están formados por nodos. (Merayo, 2020)

Los métodos basados en árboles de decisiones potencian los modelos predictivos con alta precisión, estabilidad y facilidad de interpretación.

- El algoritmo Iterative Dichotomizer 3 (ID3) fue inventado por Ross Quinlan y hasta el momento es el algoritmo más utilizado en árboles de decisión y consiste en la construcción de un árbol de decisión de arriba a abajo, de forma directa, sin hacer uso de backtracking, y basándose únicamente en los ejemplos iniciales proporcionados. Para seleccionar los criterios de decisión se considera el concepto de error de Clasificación, el índice Gini y la Entropía.

- El error de clasificación de acuerdo con la ecuación (2) es una medida de la discrepancia entre los datos reales y los predichos por el modelo. Como indicador se tiene que a menor RSS hay un buen ajuste del modelo a los datos.

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

y_i es el valor real de la variable a predecir \hat{y}_i es el valor predicho

El índice Gini de acuerdo con la ecuación (3) mide el grado de pureza de un nodo permitiendo separar mejor los ejemplos respecto a la clasificación final. Como indicador se tiene que a mayor índice de Gini menor pureza, por lo que seleccionaremos la variable con menor Gini ponderado.

$$RGINI(t) = 1 - \sum_{i=1}^n (P_i)^2 \quad (2)$$

P_i = Probabilidad de que un ejemplo sea de la clase i

La entropía de acuerdo con la ecuación (4) mide el grado de incertidumbre de una muestra. Es directamente proporcional con el índice Gini ya que ha mayor ganancia de información menor entropía ponderada de la variable. Como indicador si un nodo es puro su entropía es 0 y solo tiene observaciones de una clase, pero si la entropía es igual a 1, existe la misma frecuencia para cada una de las clases de observaciones.

$$H = - \sum_{i=1}^n P_i * \log_2 P_i \quad (3)$$

E. Análisis de resultados

Con base a los resultados obtenidos por el algoritmo el SLA óptimo a implementar en el área de estudio es el BM desarrollado en la plataforma PHYTON, como se muestra en la Figura 4. En cuanto al SLA por HL, la primera variable que el algoritmo considera para clasificar los datos es la producción acumulada en el periodo 2016-2020, al realizar la clasificación los 11 pozos analizados tuvieron una producción menor a 1252.40 barriles, este mismo caso ocurre para las variables gravedad API y ciclos en la cual todos los pozos tienen ciclos >1.5 y producen un crudo ≤ 43.05 . La siguiente variable considerada por el algoritmo es ciclo ≤ 3.5 en la cual 9 de los 11 pozos tienen ciclos mayores al establecido los dos restantes por ende tuvieron ciclos ≤ 3.5 a los cuales se los clasifica con la variable gasto total, ambos pozos tuvieron un gasto mayor a \$37176.09. Con esta clasificación se puede observar que el sistema HL no es factible a implementar en la sección 66 principalmente por su baja producción y sus altos costos de operación y mantenimiento.

En el caso del SLA por SW, de los 27 pozos que el algoritmo dispone para la prueba 8 de ellos operan con SW, la primera variable que el algoritmo considera es la producción acumulada en el periodo 2016-2020. De los ocho pozos analizados dos tuvieron una producción menor a 1252.40 barriles y seis tuvieron una producción mayor a 1252.40 barriles. La variable que considera el algoritmo para clasificar a los pozos que tuvieron una producción mayor 1252.40 barriles es gasto total ≤ 31654.59 , los seis pozos generan gastos por encima de ese valor.

Con esta clasificación se obtiene que el sistema SW no es óptimo a implementar en la sección 66 a pesar de que el 75% de los pozos que operan con este SLA tienen una producción superior a los 1252.40 barriles, la principal desventaja que presentan es que tienen gastos en operación y mantenimiento

superiores a 31654.59 dólares por lo cual no es factible para la empresa.

Finalmente, para el SLA por BM el algoritmo dispone de ocho pozos, la primera variable que el algoritmo considera es la producción acumulada en el periodo 2016-2020, siete de ocho pozos analizados tuvieron una producción mayor a 1252.40 barriles, continuando con la clasificación la siguiente variable analizada es gasto total del cual seis de los siete pozos tuvieron un gasto menor a 31654.59 dólares. Con esta clasificación el algoritmo considera que el SLA óptimo a implementar en la sección 66 es BM debido a que es el sistema con el cual se obtiene una mayor producción de petróleo y a su vez el SLA que menos gastos genera.

Para lograr estos resultados se le designó al algoritmo árbol de decisión el 15% de los datos para el entrenamiento, el porcentaje restante (85%) automáticamente los utiliza para realizar la prueba, es decir de la información de los 32 pozos que constan en la base de datos: 5 fueron usados para entrenamiento y 27 para la prueba, escogió aleatoriamente 8 pozos que operan con BM, 11 con HL y 8 con SW.

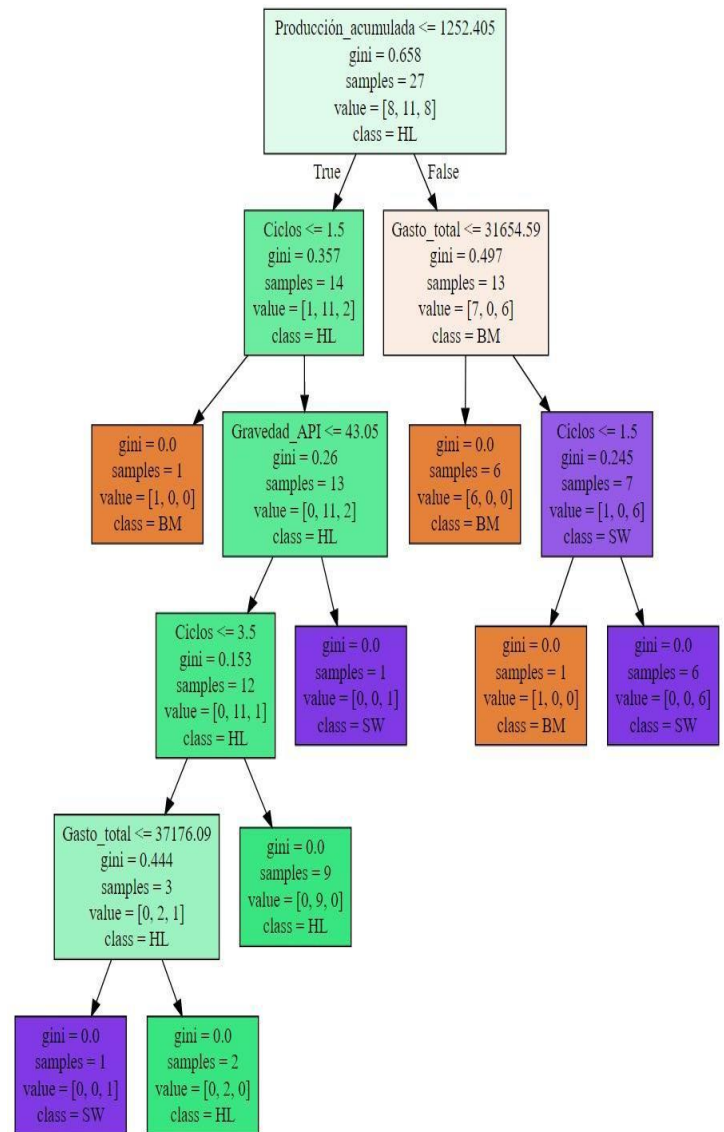


Figura 4. Resultados de las corridas utilizando el Árbol de decisión mediante ML

Para demostrar que el BM es el SLA óptimo a implementar, a continuación, se analiza los nodos del árbol obtenido: Aquellos pozos en los que se cumple la condición: producción acumulada ≤ 1252.405 barriles de petróleo producidos en el periodo 2016-2020 como muestra la Figura 5, tienden a ser clasificados hacia la izquierda (true) caso contrario los que no cumplen con dicha condición es decir que tengan una producción acumulada > 1252.405 barriles tienden a seguir hacia la derecha (false), el coeficiente de gini indica el grado de impureza que tiene cada nodo al momento de realizar la clasificación de datos, la impureza se refiere a cómo de mezcladas están las clases en cada nodo a mayor coeficiente

de gini menor será el grado de pureza, un gini de 0.658 indica un alto grado de impureza puesto que en el nodo existen tres clases (BM, HL, SW). Cabe recalcar que los 11 pozos operados con SLA por HL que el algoritmo seleccionó, obtuvieron una producción acumulada ≤ 1252.405 barriles mientras que 7 de 8 pozos que operan con SLA por BM tuvieron una producción acumulada > 1252.405 .

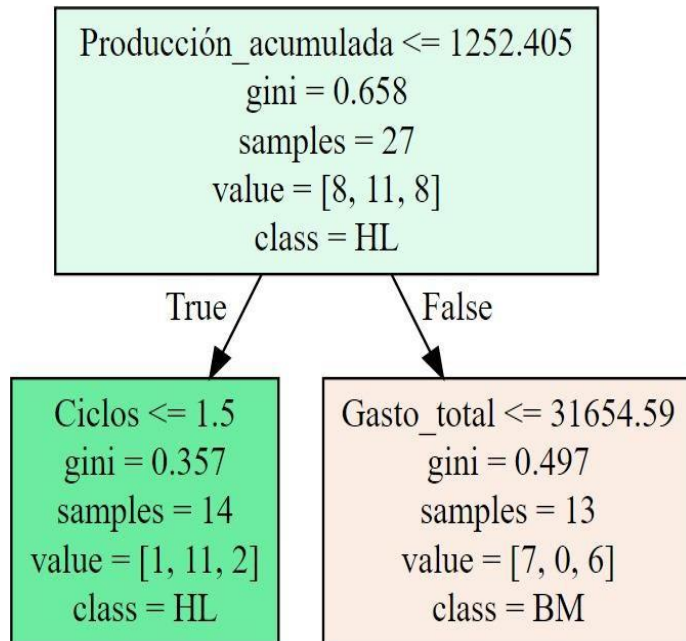


Figura 5. Nodos de producción acumulada, ciclos y gasto total

Para el caso producción acumulada ≤ 1252.405 barriles (True), se puede observar en la Figura 5 que 14 de 27 pozos cumplen con esta condición, la siguiente variable que el algoritmo utiliza para realizar la clasificación de datos es “ciclos”, la razón por la que el algoritmo escogió esta variable se debe al grado de pureza, el gini es de 0.357, al igual que en el nodo anterior se dispone de 3 clases (BM, HL, SW), se podría pensar que al tener la misma cantidad de clases el gini debería ser igual (0.658), sin embargo para la variable ciclos el coeficiente de gini ha disminuido porque en dos clases se cuenta con menor cantidad de pozos; 1 por BM y 2 por SW. Los pozos que tienden a cumplir con la condición ciclos ≤ 1.5 siguen hacia la izquierda y aquellos que tengan un ciclo > 1.5 siguen hacia la derecha.

Se puede observar en la Figura 6, que de 14 pozos uno de ellos (operado con BM) cumple con la condición ciclo ≤ 1.5 , el coeficiente de gini es 0 debido a que no existe otra clase aparte de BM, es decir que no existe impureza en el nodo, con esto el algoritmo ha clasificado el primer pozo de 27 que fueron tomados para realizar la prueba.

Los pozos que no cumplen la condición ciclo ≤ 1.5 tienden hacia la derecha, clasificándolos en 2 clases; 11 pozos con

SLA por HL y 2 por SW, el algoritmo considera otra variable para continuar con la clasificación de datos, siendo la condición gravedad api ≤ 43.05 ,

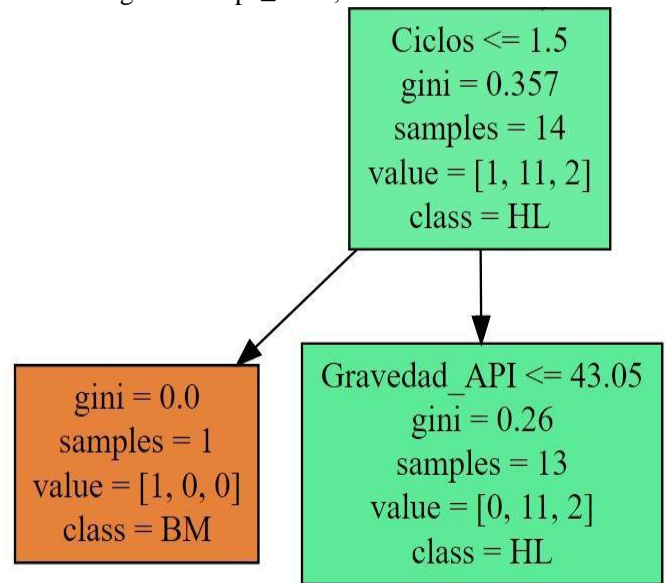


Figura 6. Nodos de ciclos y gravedad API

La siguiente variable que el algoritmo analiza es gravedad api, como muestra la Figura 7 aquellos pozos que cumplan con la condición gravedad api $\leq 43.05^\circ$ tienden a la izquierda y aquellos pozos que cumplan con gravedad api $> 43.05^\circ$ tienden hacia la derecha, En este nodo el coeficiente gini es de 0.26 ya que se dispone de 2 clases por lo tanto el grado de impureza es menor. De los 13 pozos, uno de ellos operado con SLA por SW produce un crudo con mayor gravedad a la establecida, el gini en este nodo es 0 porque no se encuentra otra clase con la cual realizar una clasificación, mientras que 12 pozos (11 operan con HL y 1 con SW) producen un crudo ≤ 43.05 .

El algoritmo analiza nuevamente la variable ciclos debido a que el gini generado es menor al que obtendría con otra variable como por ejemplo gastos, con lo cual puede clasificar los datos con un menor grado de impureza.

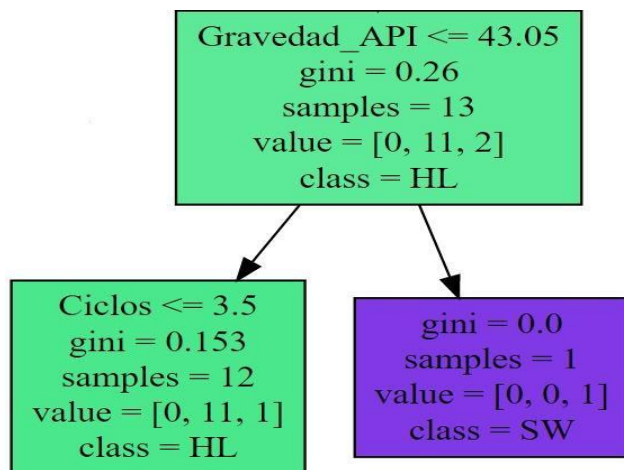


Figura 7. Nodos de gravedad API y ciclos

Los pozos que cumplen con la condición ciclos ≤ 3.5 , tienden a seguir hacia la izquierda y aquellos que tienen ciclos > 3.5 tienden hacia la derecha como se observa en la Figura 8. En este nodo el gini es de 0.153 lo cual muestra un menor grado de impureza, debido a que se tiene un pozo operado con SLA por SW y 11 con HL, cabe recalcar que el gini es 0 cuando solo quedan datos de una clase, tal caso ocurre en el nodo que se genera cuando los pozos tienen ciclos > 3.5 .

Para el caso de los pozos que cumplen la condición ciclos ≤ 3.5 , se analiza una nueva variable la cual es gasto total.

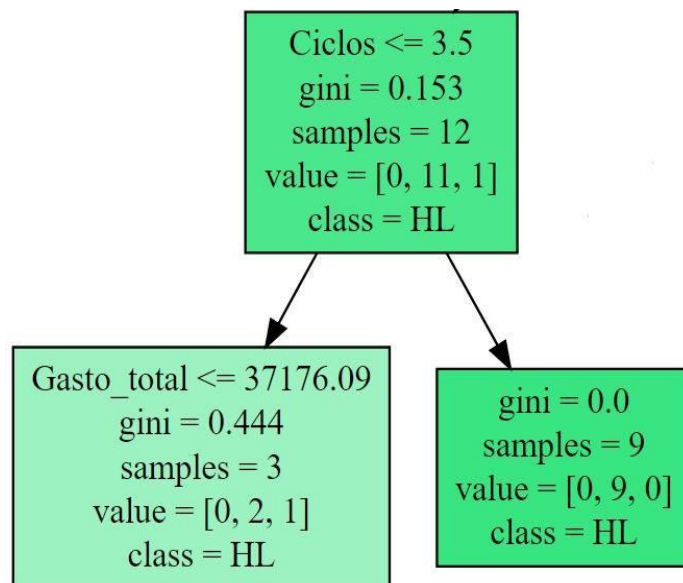


Figura 8. Nodos de ciclos y gasto total

La siguiente variable por considerar es gasto total como se muestra en la Figura 65, en la cual se tienen tres pozos, aquellos pozos que cumplen con la condición gasto total ≤ 37176.09 dólares tienden hacia la izquierda y aquellos pozos

que cumplen con gasto total > 37176.09 dólares tienden hacia la derecha, dos de tres pozos tienen un gasto mayor al especificado, ambos pozos operan con el sistema HL, mientras que el pozo que tiene un gasto menor opera con SLA por SW. En ambos nodos el gini es cero debido a que no existe otra clase con la cual se pueda realizar una siguiente clasificación.

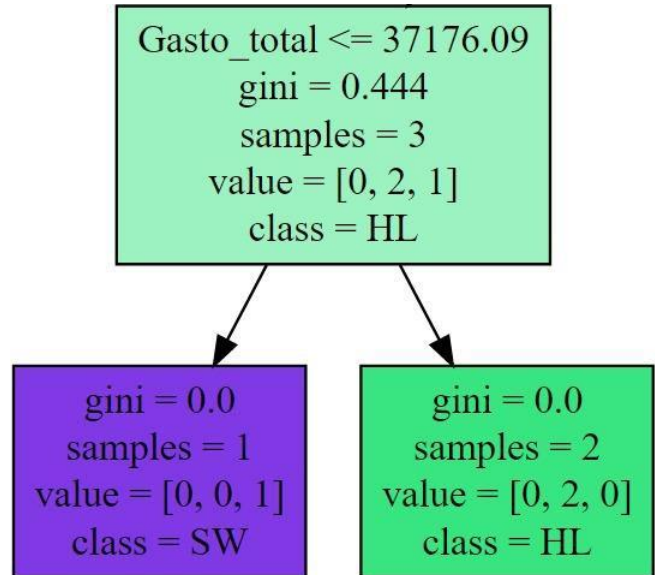


Figura 9. Nodo de gasto total

Los pozos que tienen una producción acumulada > 1252.405 barriles tienden hacia la derecha como se muestra en la Figura 10, en este nodo se puede observar que 13 pozos cumplieron con esta característica, de los cuales 7 operan con SLA por BM y 6 operan con el SLA por SW [7,0,6].

En la Figura 10 se observa que el algoritmo utiliza la condición gasto total ≤ 31654.59 para clasificar los datos, aquellos pozos que cumplen dicha condición tienden hacia la izquierda y los que generan un gasto superior a ese valor tienden hacia la derecha. De los siete pozos que operan con BM, seis tuvieron un gasto total ≤ 31654.59 en este nodo el gini es de cero debido a que es la única clase presente. Los Pozos que generan un gasto mayor al de la condición son; seis por SW y uno opera con SLA por BM.

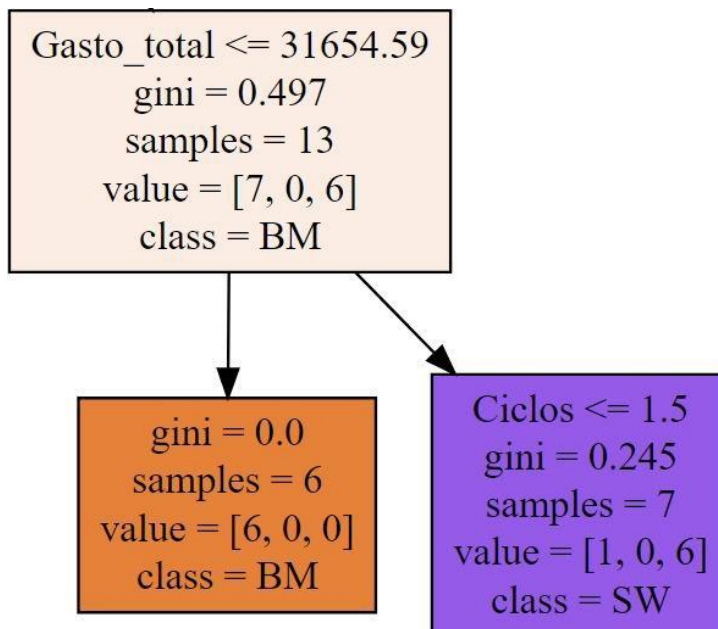


Figura 10. Nodos de gasto total y ciclos

En la Figura 11 el algoritmo considera la variable ciclos para realizar la siguiente clasificación, para este caso separa los pozos que tienen ciclos ≤ 1.5 de aquellos que tienen ciclos mayores. Se dispone de siete pozos, al realizar la clasificación seis pozos que operan con SLA por SW tiene ciclos > 1.5 , mientras que un pozo que opera con BM tiene ciclo ≤ 1.5 .

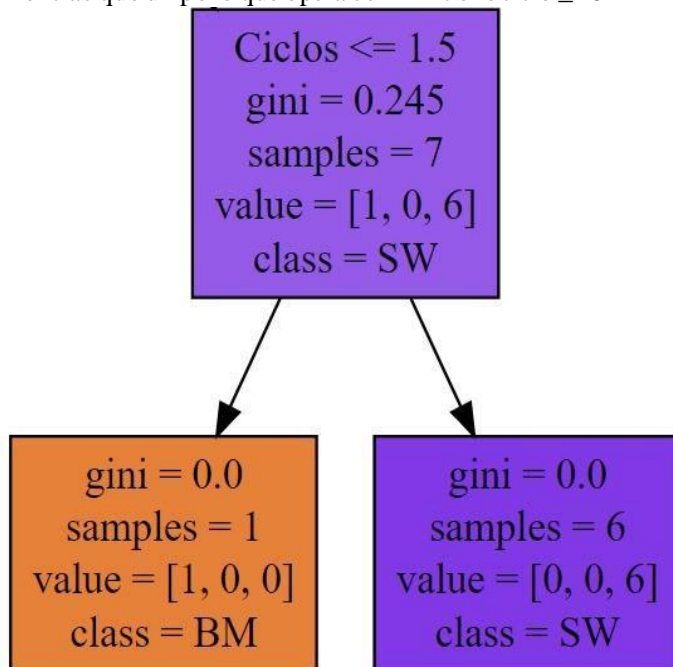


Figura 11. Nodo de ciclos

En la Figura 12 se observa la clasificación de datos del algoritmo árbol de decisión utilizando dos variables; producción acumulada y gasto total, da como resultado que el SLA óptimo a implementar en toda la sección 66 es BM. Los pozos que operan con BM obtuvieron una mejor tasa de producción en el periodo 2016-2020 en comparación a los pozos que producen con otros SLA, y a su vez generaron menor gasto de operación y mantenimiento debido a que las intervenciones que se realizan en los pozos son de bajo costo, además se economiza dinero en combustible, grasas, lubricantes y repuestos.

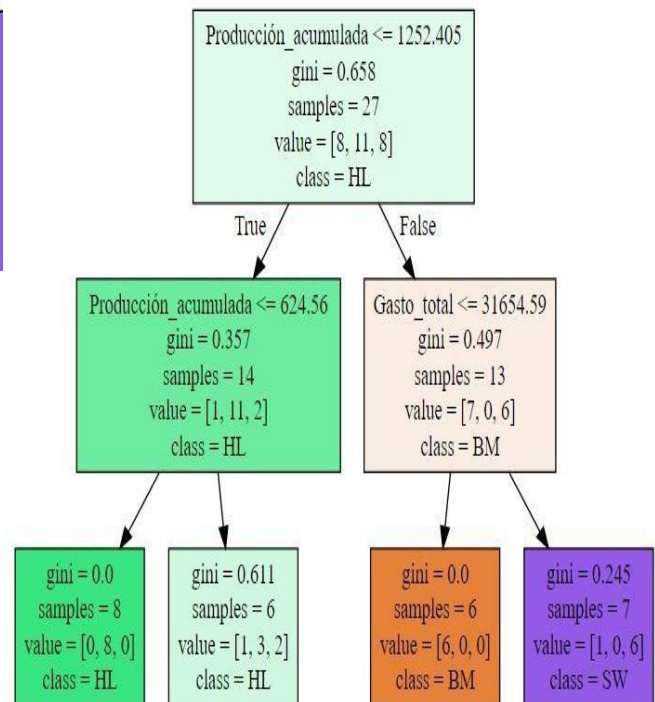


Figura 12. Árbol de decisión con dos variables

F. Conclusiones

- En la sección 66 hay un total de 60 pozos, la producción de petróleo se obtiene de 32 pozos mediante tres tipos de sistemas de levantamiento artificial mismos que fueron utilizados por el algoritmo a fin de seleccionar un único sistema, estos son: bombeo mecánico, swab y herramienta local, los 28 pozos restantes no fueron considerados debido a que presentan problemas, razón por la cual se encuentran cerrados y uno de ellos es pozos inyector.
- Los sistemas de levantamiento artificial con el que operan los pozos de la sección 66 presentan sus respectivas ventajas y desventajas que los diferencian uno de otro. El bombeo mecánico es un sistema que ofrece mejores resultados debido a puede ser utilizado prácticamente durante toda la vida productiva del pozo, además los costos de operaciones son bajos, por el contrario, los sistemas herramienta local y

swab presenta altos costos de operación y mantenimiento, este último al evaluar el pozo requiere bajar tubería lo cual demanda un gasto extra, siendo el sistema de levantamiento artificial más costoso.

- El uso de un algoritmo supervisado de machine learning basado en árbol de decisión permite obtener mejor predicción en comparación a modelos de regresión lineal y regresión polinómica. El objetivo propuesto se cumple realizando una clasificación de datos con un análisis de las variables disponibles en el dataset, dicha tarea la resuelven los algoritmos de clasificación puesto que estos predicen resultados de acuerdo con el entrenamiento y la variable dependiente que ha recibido.

- La predicción del algoritmo árbol de decisión da como resultado que el sistema de levantamiento artificial óptimo a implementar en toda la sección 66 es bombeo mecánico, porque los pozos que operan con este sistema obtuvieron una mayor tasa de producción durante el periodo 2016-2020 en comparación a los pozos que producen con otros sistemas de levantamiento artificial, y a su vez generaron menor gasto de operación y mantenimiento debido a que las intervenciones que se realizan en los pozos son de bajo costo, además se economiza dinero en combustible, grasas, lubricantes y repuestos.

ACKNOWLEDGMENT

Especial agradecimientos a la Universidad Estatal Península de Santa Elena.

REFERENCES

- [1] 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.
- [2] M. R. Khan, "2 Spe-194713-MS," 2019.
- [3] M. Pennel, J. Hsiung, and V. B. Putcha, "Detecting failures and optimizing performance in artificial lift using machine learning models," *SPE West. Reg. Meet. Proc.*, vol. 2018-April, 2018.
- [4] T. Ounsakul, A. Rittirong, T. Kreethapon, W. Toempromraj, and K. Wejwittayaklung, "SPE-196470-MS Data-Driven Diagnosis for Artificial Lift Pump's Failures Introduction Datasets for Pump Failure Diagnostics," 2019.
- [5] C. Carpenter, "Dynamometer-Card Classification Uses Machine Learning," *J. Pet. Technol.*, vol. 72, no. 03, pp. 52–53, 2020.
- [6] S. A. Sharaf, P. Bangert, M. Fardan, K. Alqassab, M. Abubakr, and M. Ahmed, "Beam pump dynamometer card classification using machine learning," *SPE Middle East Oil Gas Show Conf. MEOS, Proc.*, vol. 2019-March, 2019.
- [7] N. J. Van Rensburg, V. P. Oil, and G. U. Solutions, "Why Keep Frac Pumps," no. May, 2019.
- [8] García, P., & Villón, M. (2014). Repositorio Dspace UPSE. Obtenido de <https://repositorio.upse.edu.ec/xmlui/handle/46000/1666>
- [9] Alonso, R. (8 de Abril de 2020). Hard Zone. Obtenido de <https://hardzone.es/tutoriales/rendimiento/diferencias-ia-deep-machine-learning/>
- [10] Alandete, D. (27 de Octubre de 2011). El País. Obtenido de John McCarthy, el arranque de la inteligencia artificial: https://elpais.com/diario/2011/10/27/necrologicas/1319666402_850215.html
- [11] Sanseviero, O. (30 de Enero de 2018). AI Learners. Obtenido de <https://medium.com/ai-learners/ai-en-3-minutos-tipos-de-machine-learning-945b708ac78>
- [12] Merayo, David, Alvaro Rodríguez-Prieto, and Ana M. Camacho. 2021. "Topological Optimization of Artificial Neural Networks to Estimate Mechanical Properties in Metal Forming Using Machine Learning" *Metals* 11, no. 8: 1289. <https://doi.org/10.3390/met11081289>