

Screening Enhanced Oil Recovery methods using the machine-learning algorithm K Nearest Neighbors (KNN)

Carrión Maldonado Freddy Paúl, Ing.¹, Lliguizaca Dávila Jorge Rodrigo, M.Sc.¹, Perero Macías Damian Leonardo, Ing.¹
Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias de la Tierra, Campus Gustavo Galindo Km 30.5
Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador, fpcarrio@espol.edu.ec, jorollig@espol.edu.ec,
dlperero@espol.edu.ec

Abstract– The process of selecting enhanced oil recovery methods, known as EOR (Enhanced Oil Recovery) methods, has been a great challenge within the oil and gas industry because it depends on the characteristics of the reservoir and economic aspects. This has usually been done through the use of comparative tables, where the application of each of these methods is determined by specific ranges of reservoir parameters. However, this selection methodology may generate biased results due to its qualitative nature and interpretation depending on the criteria of the professional in charge of the process, therefore, it is pertinent to seek the implementation of techniques that provide better results.

Therefore, this article focuses on improving the accuracy of the selection of these methods, implementing data-driven decisions through a machine-learning model, specifically the Nearest K model, whose objective is to predict which enhanced oil recovery method is the most adequate to implement in an oil well located in the Amazon region of Ecuador, whose reservoir contains heavy oil. It is worth mentioning that the algorithm was trained with more than 200 enhanced oil recovery projects from several oil-producing countries, where the input data of the model are seven reservoir parameters, which are porosity, permeability, depth, temperature, API gravity, viscosity, and oil saturation whereas the output parameters of the model are 5 enhanced recovery methods, which are steam injection, CO₂ injection, hydrocarbon injection, polymer injection, and in-situ combustion.

The Model was successfully trained because its metrics yielded acceptable values, as its accuracy was about 84%, the error rate equal to 0.16, and as a result, it predicted that the most suitable EOR method to implement into this oil well is the steam injection method.

Keywords: EOR methods selection, machine-learning, K-nearest neighbors model.

Digital Object Identifier (DOI):
<http://dx.doi.org/10.18687/LACCEI2020.1.1.564>
ISBN: 978-958-52071-4-1 ISSN: 2414-6390

Selección de métodos de recuperación mejorada de petróleo utilizando el algoritmo de machine-learning de los K vecinos más cercanos (KNN)

Carrión Maldonado Freddy Paúl, Ing.¹, Lliguizaca Dávila Jorge Rodrigo, M.Sc.¹, Perero Macías Damian Leonardo, Ing.¹

Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias de la Tierra, Campus Gustavo Galindo Km 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador, fpcarrio@espol.edu.ec, jorollig@espol.edu.ec, dlperero@espol.edu.ec

Resumen. – *El proceso de selección de métodos de recuperación mejorada de petróleo, conocidos como métodos EOR (Enhanced Oil Recovery), ha representado un gran desafío dentro de la industria del petróleo y gas, debido a que el mismo depende de las características del reservorio y de aspectos económicos. Usualmente se lo ha llevado a cabo mediante el uso de tablas comparativas, en donde la aplicación de cada uno de estos métodos está determinada por rangos específicos de los parámetros del yacimiento. Sin embargo, esta metodología de selección puede generar resultados sesgados por su carácter cualitativo e interpretación dependiente del criterio del profesional a cargo del proceso, por lo tanto, es pertinente buscar la implementación de técnicas que proporcionen mejores resultados.*

Por consiguiente, el presente artículo se enfoca en mejorar la precisión de la selección de estos métodos, implementando tomas de decisiones basadas en datos a través de un modelo de machine-learning, específicamente el modelo de los K vecinos más cercanos, cuyo objetivo es predecir que método de recuperación mejorada es el más adecuado a implementar en un pozo de petróleo ubicado en la región amazónica del Ecuador, cuyo reservorio contiene crudo pesado. Cabe mencionar, que el algoritmo fue entrenado con más de 200 proyectos de recuperación mejorada de varios países petroleros, en donde los datos de entrada del modelo son siete parámetros de yacimiento, los cuales son porosidad, permeabilidad, profundidad, temperatura, gravedad API, viscosidad, y saturación de petróleo mientras que los parámetros de salida del modelo son 5 métodos de recuperación mejorada, los cuales son inyección de vapor, inyección de CO₂, inyección de hidrocarburos, inyección de polímeros y combustión in situ.

El Modelo fue entrenado con éxito, debido a que sus métricas arrojaron valores aceptables, ya que su precisión fue de alrededor del 84%, la tasa de error igual a 0.16 y como resultado, predijo que el método EOR más adecuado a implementar en el pozo de petróleo es el método de inyección de vapor.

Palabras clave: Selección de métodos EOR, machine-learning, Modelo de los K vecinos más cercanos.

Abstract- *The process of selecting enhanced oil recovery methods, known as EOR (Enhanced Oil Recovery) methods, has been a great challenge within the oil and gas industry because it depends on the characteristics of the reservoir and economic aspects. This has usually been done through the use of comparative tables, where the application of each of these methods is determined by specific ranges of reservoir parameters. However, this selection methodology may generate biased results due to its qualitative nature and interpretation depending on the criteria of the professional in charge*

of the process, therefore, it is pertinent to seek the implementation of techniques that provide better results.

Therefore, this article focuses on improving the accuracy of the selection of these methods, implementing data-driven decisions through a machine-learning model, specifically the Nearest K model, whose objective is to predict which enhanced oil recovery method is the most adequate to implement in an oil well located in the Amazon region of Ecuador, whose reservoir contains heavy oil. It is worth mentioning that the algorithm was trained with more than 200 enhanced oil recovery projects from several oil-producing countries, where the input data of the model are seven reservoir parameters, which are porosity, permeability, depth, temperature, API gravity, viscosity, and oil saturation whereas the output parameters of the model are 5 enhanced recovery methods, which are steam injection, CO₂ injection, hydrocarbon injection, polymer injection, and in-situ combustion.

The Model was successfully trained because its metrics yielded acceptable values, as its accuracy was about 84%, the error rate equal to 0.16, and as a result, it predicted that the most suitable EOR method to implement into this oil well is the steam injection method.

Keywords: EOR methods selection, machine-learning, K-nearest neighbors model.

I. INTRODUCCIÓN

Los avances tecnológicos han permitido desarrollar nuevas técnicas para la producción de hidrocarburos y la extensión de su vida productiva, entre ellos destacan los métodos de recuperación mejorada de petróleo los cuales han dado excelentes resultados en cuanto al considerable aumento del factor de recobro, sin embargo presenta limitaciones económicas y la selección optima de un método de recuperación mejorada es una parte crítica para su implementación[1].

Estos métodos han sido estudiados durante décadas, pero su implementación en campo ha sido bastante escasa. Los autores en [2] concluyen enfatizando la importancia de la planeación de estos proyectos donde la selección de la técnica más adecuada es de suma importancia para asegurar el éxito del proyecto.

Es importante entender mejor el proceso y mejorar los criterios de selección del tipo de método de acuerdo con el yacimiento y sus fluidos. Distintos autores han propuesto criterios y métodos de selección para este tipo de técnicas, entre ellos, en el estudio presentado por [3] se ha presentado un

Digital Object Identifier (DOI):

<http://dx.doi.org/10.18687/LACCEI2020.1.1.566>

ISBN: 978-958-52071-4-1 ISSN: 2414-6390

proceso estadístico a través de un sistema experto de selección para evaluar la aplicación de un método de EOR en un reservorio. Lo cual demuestra la importancia de la recopilación de información y la aplicación de herramientas matemáticas que permitan generar sistemas de soporte de decisiones.

El algoritmo de KNN (vecinos más cercanos) es uno de los más comúnmente usados en una amplia variedad de campos [4], ha sido usado en los campos del área de química, medicina, economía, energía, geología, ingeniería en general; y para el actual estudio en el área petrolera específicamente en lo que respecta a la aplicación de métodos de recuperación mejorada de petróleo.

El modelo matemático KNN (vecinos más cercanos) es un algoritmo basado en el aprendizaje y estimación de un conjunto de datos de entrenamiento prototipo que sirve para estimar una función de densidad, para la predicción por cada clase ingresada en el método; mismo que sigue el siguiente proceso extraído de [5]

Fix and Hodges (1952) consideraron al método no paramétrico KNN para trabajar con $K=1$ es decir para una regla paramétrica lineal. Sin embargo en investigaciones más recientes con tablas y gráficos de rendimiento para una variedad de separaciones medias y pequeños tamaños de muestra sugieren que su desempeño no es muy inferior al de la regla paramétrica lineal cuando se cumplen los supuestos en los que se base este último.[6]

El ser de los algoritmos más usados, se lo debe a sus características principales: Su simplicidad conceptual y su buena caracterización de resultados tanto como hayan sido bien seleccionados la base de datos para el entrenamiento del método.[7]

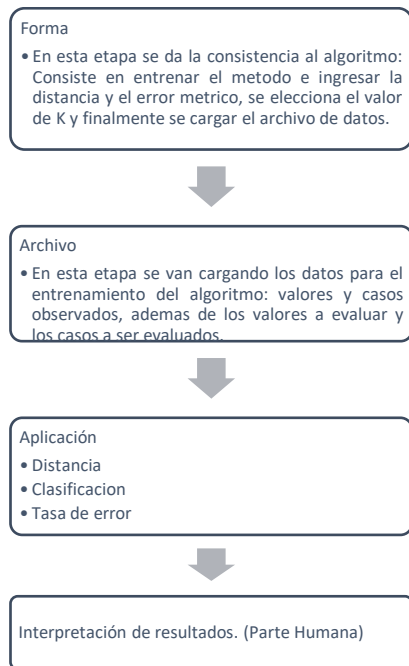


Fig 1. Esquema del algoritmo de los K vecinos más cercanos [5]

Además, es pertinente acotar que el algoritmo matemático KNN (vecinos más cercanos) tiende a funcionar mejor con conjuntos de datos pequeños que no tengan muchas características (variables independientes) [8] debido a que el algoritmo analiza todos los puntos de datos para cada clasificación, por lo que si se lo entrena con grandes cantidades de datos, los resultados podrían tener una tasa de error alta.

Por tanto, el presente estudio pretende demostrar la aplicación y utilidad del algoritmo matemático de los K vecinos más cercanos mediante el uso del lenguaje de programación Python para seleccionar el método de EOR más apropiado dadas las características del pozo seleccionado para el caso de estudio.

II. METODOLOGÍA

La selección del método más adecuado de recuperación mejorada (método EOR) para ser implementado en el pozo de petróleo del caso de estudio, se basa en la aplicación del algoritmo de machine-learning de los K vecinos más cercanos, el cual es utilizado para predecir variables categóricas. Debido a que los parámetros de salida del algoritmo son 5 métodos de recuperación mejorada, por lo tanto, este algoritmo es apropiado para la predicción y el tipo de datos de salida, por lo cual fue diseñado y entrenado para predecir variables multiclasas.

El entrenamiento, evaluación y ejecución del algoritmo se lo llevó a cabo mediante la aplicación del programa Python a través de su librería scikit-learn, la cual posee códigos de programación que permiten la ejecución de una amplia gama de los algoritmos de machine-learning entre los que se encuentra el algoritmo utilizado.

El algoritmo en mención fue entrenado con más de 200 proyectos de recuperación mejorada ejecutados en varios países petroleros, de los cuales los parámetros de entrada para el entrenamientos del algoritmos son 7 siete parámetros de yacimiento, estos son: porosidad, permeabilidad, profundidad, temperatura, gravedad API, viscosidad, y saturación del petróleo, mientras que como parámetros de salida fueron escogidos 5 métodos de recuperación mejorada, los cuales son inyección de vapor, inyección de CO₂, inyección de hidrocarburos, inyección de polímeros, y combustión in situ.

A. Caso de estudio

El caso de estudio de esta investigación es un pozo de petróleo al cual denominaremos pozo ESPOL 01, el cual se encuentra ubicado en el noreste de la república del Ecuador, cercano a sus fronteras con las repúblicas de Colombia y Perú, específicamente se encuentra en la región amazónica del Ecuador como se muestra en la Fig. 2, en la cuenca Oriente. El reservorio al que pertenece dicho pozo tiene una gravedad API muy baja, lo cual hace referencia que el crudo almacenado es considerado pesado.

Cabe mencionar que este pozo fue perforado hace aproximadamente 20 años por otra compañía, la cual en ese entonces estuvo a cargo de la concesión de dicho campo. Sin

embargo, debido a la baja rentabilidad del proyecto, en especial contrastado por la baja gravedad API de los reservorios del campo, conllevó posteriormente a que la compañía finiquite su contrato con el estado ecuatoriano.

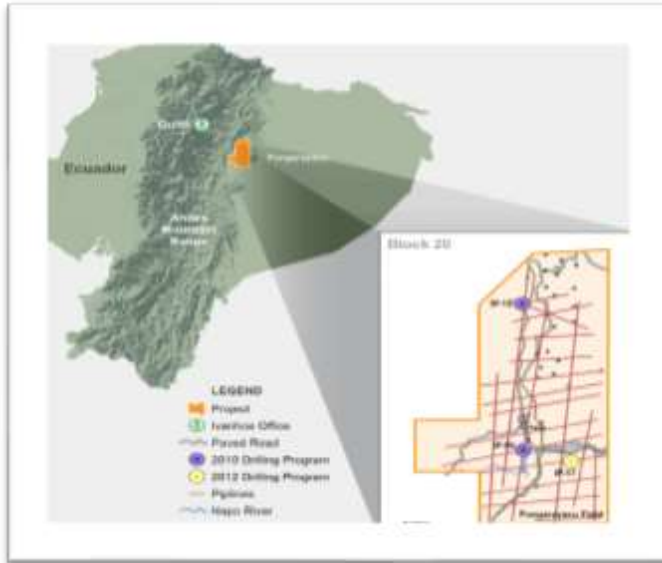


Fig 2. Localización del campo de petróleo del caso de estudio

Los principales parámetros del yacimiento al que pertenece el pozo ESPOL 01 fueron proporcionados por la compañía que estuvo a cargo de la concesión del campo de petróleo [9], los cuales son mostrados en la tabla I, en donde se encuentran los valores de porosidad, permeabilidad, profundidad, gravedad API, viscosidad, temperatura, y saturación del petróleo.

TABLA I
PARÁMETROS DEL YACIMIENTO

Parámetros de yacimiento	Unidad	Valor
Porosidad	%	29
Permeabilidad	md	4689
Profundidad	ft	1200
Gravedad API	Adimensional	8
Viscosidad	cp	490058
Temperatura	F	150
Saturación inicial del Petróleo	%	38.4

B. Definición del problema

El principal desafío que conlleva la selección de métodos de recuperación mejorada 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 un algoritmo de machine-learning para la realización de dicha selección, con el objetivo de demostrar que este desafío puede ser superado mediante la implementación de la técnica computacional en mención.

Por otro lado, es pertinente describir las principales características de algoritmo que se ha utilizado en este estudio. Como ya se mencionó previamente, el modelo de los K vecinos más cercanos se utilizó para la realización de la selección de la técnica más adecuada a implementar en el pozo de petróleo del caso de estudio. Por lo tanto, de la Fig. 3 [10] se puede visualizar mediante un ejemplo como es el funcionamiento de dicho algoritmo, de donde se puede apreciar que el resultado esperado (representado por el círculo verde) está entre dos clases que están representadas por los cuadrados azules y los triángulos rojos, sin embargo, el círculo verde (salida prevista) está más cerca de los triángulos rojos, por lo tanto, se podría estimar que el mismo pertenece a la clase representada por los triángulos rojos, debido a que este algoritmo tiene como objetivo calcular la distancia mínima del resultado esperado a la clase más cercana según la distancia euclidiana [10], la cual puede ser determinada con la ecuación que se muestra a continuación (1):

$$D = \sqrt{\sum_{i=1}^K (x_i - y_i)^2} \quad (1)$$

Donde:

D : Distancia Euclidiana

k : Numero de clases o categorías

x_i, y_i : puntos para calcular la distancia mínima de acuerdo con la distancia Euclidiana

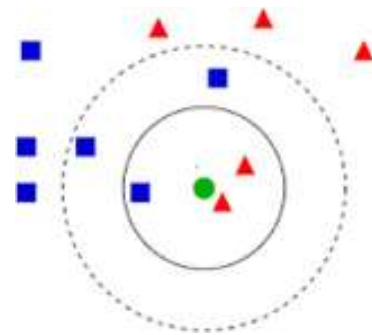


Fig 3. Ilustración del algoritmo de los K vecinos más cercanos

C. Procedimiento

Paso 1. Importar a Python en formato CSV la data proveniente de más de 200 proyectos exitosos de recuperación mejorada de

petróleo [11] realizados en varios países alrededor del mundo (ver TABLA II).

Paso 2. Visualizar como está distribuido el conjunto de datos y posteriormente realizar el preprocesamiento de datos, en donde se busca eliminar datos nulos (Ver Fig. 4), normalizar los valores de todos los datos, así como etiquetar los valores de las variables categóricas (ver TABLA III), debido a que los 5 métodos de recuperación mejorada de petróleo que constan como los valores de salida del algoritmo tienen que ser etiquetados para que Python los pueda manejar adecuadamente.

Paso 3. Separar el conjunto de datos en datos de prueba y datos de entrenamiento, en donde para este proyecto, se ha implementado la distribución standard de la cantidad de datos de prueba y entrenamiento que se utilizan en los algoritmos de machine-learning, la cual consiste en separar el 30% de los datos como datos de prueba y el 70% de los datos restantes como datos de entrenamiento. Esta separación de los datos fue implementada utilizando la librería de Python scikit-learn.

Paso 4. Posteriormente se procede a entrenar el algoritmo, utilizando nuevamente la librería scikit-learn, debido a que la misma contiene un código de programación mediante el cual el algoritmo se lo entrena directamente.

Paso 5. Después que el algoritmo se ha entrenado, se evalúa su comportamiento, mediante el cálculo de sus métricas más importantes tales como: la precisión, la tasa de error, la exhaustividad, y F1 score.

Paso 6. Con el objetivo de visualizar el comportamiento del algoritmo, son construidos 2 gráficos los cuales son precisión vs el número de K vecinos y tasa error vs el número de K vecinos, esto con el fin de determinar qué valor de K, dará como resultado la precisión más alta y la tasa de error más baja, que es lo que busca el algoritmo para que su entrenamiento sea exitoso, y por consiguiente esté listo para realizar la evaluación de los métodos de recuperación mejora de petróleo.

TABLA II
PROYECTOS EOR POR PAÍS

Proyectos EOR	País
USA	140
Canada	38
Brasil	8
Venezuela	1
Noruega	5
Alemania	10
Total	202

De la tabla II, se puede apreciar la distribución por país de los proyectos de recuperación mejorada de petróleo que fueron escogidos como el conjunto de datos para entrenar y evaluar el algoritmo, de donde Estados Unidos se destaca como el país con mayor cantidad de este tipo de proyectos, debido a que cuenta con 140.

TABLA III
ETIQUETA DE LOS PARÁMETROS DE SALIDA

Método EOR	Etiqueta
Inyección de CO2	0
Inyección de hidrocarburos	1
Combustión	2
Inyección de polímeros	3
Inyección de vapor	4

De la tabla III, se puede visualizar los valores etiquetados con los que se identificó a cada uno de los 5 métodos EOR evaluados en este estudio para que puedan ser manejados en el programa Python, los cuales fueron representados con valores del 0 al 4, basándose en orden de alfabético.

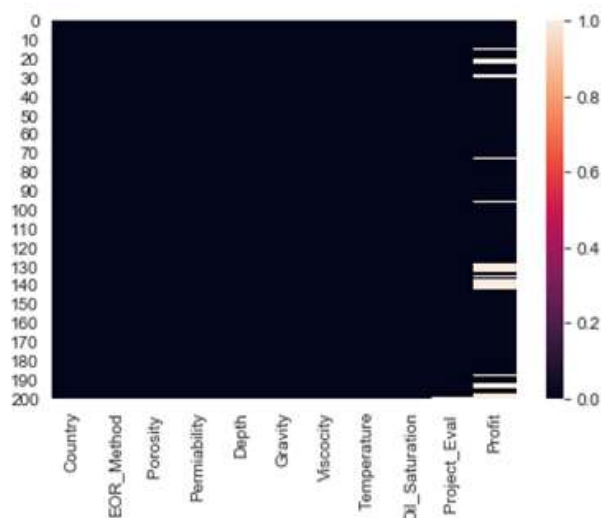


Fig 4. Mapa de calor para identificar datos nulos

En la Fig. 4 se puede apreciar mediante un mapa de calor los datos nulos de los parámetros de yacimientos del conjunto de datos. Sin embargo, se puede identificar que solo hay datos nulos en el parámetro Profit, el cual no afecta a la evaluación del algoritmo, debido a que no fue seleccionado como variable independiente del modelo matemático.

III. ANÁLISIS Y RESULTADOS

A. Evaluación del desempeño del algoritmo de los K vecinos más cercanos

El rendimiento del algoritmo de los K vecinos más cercanos fue evaluado mediante el cálculo de las métricas utilizadas para algoritmos cuyos resultados sean variables categóricas, tal como es el caso del usado en este artículo. Por lo tanto, las métricas implementadas fueron precisión, exhaustividad, F1 score, y tasa de error.

Por otro lado, dicha evaluación fue realizada tomando en cuenta 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), es decir el cálculo de las métricas mencionadas en el párrafo anterior fue realizada para cada valor de K , lo cual se lo puede corroborar y visualizar en la tabla IV.

De acuerdo con los resultados de la tabla IV, cuando el valor de K es 2, es decir, cuando se toman solo 2 valores alrededor del valor a evaluar, el algoritmo presenta su mejor rendimiento, debido a que a estas condiciones su precisión es del 84%, el cual es un indicador aceptable y además es el más alto en comparación a los calculados para el resto de los valores de K . Por otro lado, la tasa de error a esta condición es 0.052, la cual es bastante aceptable, ya que el objetivo es que la tasa de error sea lo más baja posible, además de que es la más baja en comparación a las calculados para el resto de los valores de K .

Finalmente, los resultados de exhaustividad y F1 score a la condición de K igual 2 también son aceptables en comparación a los calculados para el resto de los valores de K . Por lo tanto, se puede deducir, que el algoritmo esta correctamente entrenado y listo para realizar la evaluación de los métodos de recuperación mejorada del petróleo, cuando el valor de los K vecinos es igual 2.

TABLA IV
EVALUACIÓN DEL RENDIMIENTO DEL ALGORITMO
DE LOS K VECINOS MÁS CERCANOS

K vecinos	Precisión	Exhaustividad	F1 Score	Tasa de error
2	0.84	0.7	0.69	0.052
5	0.82	0.77	0.7	0.072
10	0.8	0.57	0.55	0.079
15	0.79	0.48	0.44	0.082
17	0.79	0.48	0.44	0.072
20	0.77	0.46	0.41	0.069
25	0.75	0.42	0.38	0.072
28	0.74	0.32	0.34	0.078
30	0.74	0.32	0.34	0.078

B. Visualización gráfica del desempeño del algoritmo de los K vecinos más cercanos

Con el objetivo de visualizar que tan optimo ha sido el desempeño del algoritmo, fueron graficadas 2 curvas, la Fig. 5 la cual muestra la precisión versus el número de los K vecinos y la Fig. 6 que muestra la tasa de error versus el número de los K vecinos.

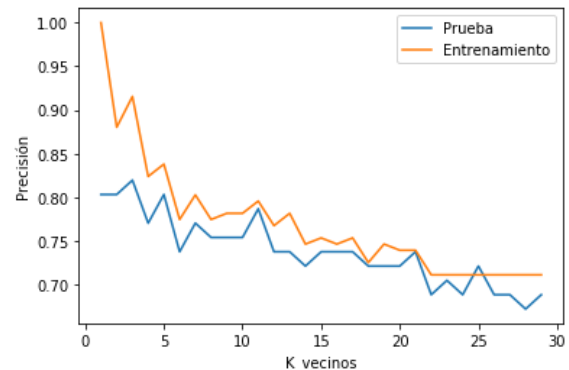


Fig 5. Curvas de precisión

Tanto en la Fig. 5 como en la Fig. 6 fueron graficadas las curvas utilizando los datos de prueba y entrenamiento, con el fin de determinar qué tan similar es el desempeño del algoritmo para cada conjunto de datos.

De la Fig. 5 se puede visualizar en la parte izquierda del gráfico, que cuando el valor de K es igual 2, la precisión del algoritmo utilizando el conjunto de datos de prueba (grafico azul) tiene su valor más alto, el cual es aproximadamente 84 %, mientras que en la parte derecha del grafico cuando K es igual a 29, la precisión del algoritmo utilizando el conjunto de datos de prueba tiene su valor más bajo, aproximadamente 69%, es decir, se puede identificar que la tendencia de la precisión es decreciente a medida que los valores de los K vecinos aumentan. Además, se puede visualizar que el comportamiento de la curva utilizando los datos de prueba es muy similar al de la curva que utiliza los datos de entrenamiento, debido a que ambas tienen la misma tendencia.

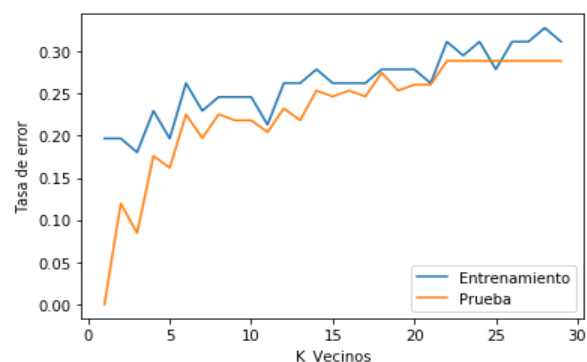


Fig 6. Curvas de la tasa de error

De la Fig. 6 se puede visualizar en la parte izquierda del gráfico, que cuando el valor de K es igual 2, la tasa de error del algoritmo utilizando el conjunto de datos de prueba (gráfico azul) tiene su valor más bajo, el cual es aproximadamente 0.052, mientras que en la parte derecha del gráfico cuando K es igual a 29, la tasa de error del algoritmo utilizando el conjunto de datos de prueba tiene su valor más alto, aproximadamente 0.079, es decir, se puede identificar que la tendencia de la tasa de error es creciente a medida que los valores de los K vecinos aumentan. Además, se puede visualizar que el comportamiento de la curva utilizando los datos de prueba es muy similar al de la curva que utiliza los datos de entrenamiento, debido a que ambas tienen la misma tendencia.

C. Selección del método de recuperación mejorada de petróleo

Una vez que el algoritmo fue correctamente entrenado, se procedió a realizar la evaluación de los métodos de recuperación mejorada, con el fin de identificar cuál de los 5 métodos que fueron analizados es el más adecuado para implementarlo en el pozo de petróleo del caso de estudio. Cabe resaltar que, para realizar tal evaluación, los datos de entrada del algoritmo fueron los parámetros de yacimiento del pozo en mención, los cuales fueron descritos en la tabla I.

La evaluación dio como resultado, que el método EOR más adecuado a implementar en este pozo de petróleo es el método de inyección de vapor. En la tabla V, se puede visualizar un resumen de los principales resultados obtenidos durante el entrenamiento y evaluación del algoritmo KNN (vecinos más cercanos).

TABLA V
RESULTADOS DEL ENTRENAMIENTO Y
EVALUACIÓN DEL ALGORITMO DE LOS K VECINOS
MAS CERCANOS KNN

K vecinos	Precisión	Tasa de error	Método EOR seleccionado
2	0.84	0.16	Inyección de vapor

IV. CONCLUSIONES

La evaluación de los métodos EOR se llevó a cabo con éxito mediante la aplicación del algoritmo de machine-learning de los K vecinos más cercanos (KNN). Este algoritmo se utilizó para predecir cual método EOR es el más adecuado para implementarlo en el pozo de petróleo del caso de estudio. Los métodos tomados en cuenta para estudio fueron: combustión in situ, inyección de CO₂, inyección de hidrocarburos, inyección de polímeros e inyección de vapor, mientras que los parámetros de entrada del algoritmo para esta evaluación fueron siete parámetros del yacimiento, los cuales son: porosidad,

permeabilidad, profundidad, saturación de petróleo, temperatura, gravedad API y viscosidad.

El algoritmo fue entrenado y probado para realizar la evaluación utilizando el software Python, así como sus paquetes para implementar algoritmos de aprendizaje automático tales como pandas, Numpy, matplotlib, y Scikit-learn. El desempeño del algoritmo fue bastante aceptable ya que alcanzó buenos valores de precisión, tasa de pérdida, exhaustividad y F1 score y finalmente, predijo que el método EOR a implementar en el pozo de petróleo del caso de estudio debería ser el método de inyección de vapor.

El algoritmo demuestra su capacidad para predecir el método de recuperación mejorada a seleccionarse, tomando en cuenta que la cantidad de datos disponibles no era muy extensa. Para futuras aplicaciones de otros tipos de algoritmos de machine-learning es importante el disponer de una extensa base de datos.

V. REFERENCIAS

- [1] J. M. Sharp, "the Potential of Enhanced Oil Recovery Processes," *Soc. Pet. Eng. - Fall Meet. Soc. Pet. Eng. AIME, FM 1975*, 1975, doi: 10.2523/5557-ms.
- [2] J. R. Hite, S. M. Avasthi, and P. L. Bondor, "Planning EOR projects," *SPE Int. Pet. Conf. Mex. - Proc.*, pp. 497–504, 2004, doi: 10.2523/92006-ms.
- [3] J. Moreno, O. Gurpinar, Y. Liu, A. Al-Kinani, and N. Cakir, "EOR advisor system: A comprehensive approach to EOR selection," *Soc. Pet. Eng. - Int. Pet. Technol. Conf. 2014, IPTC 2014 - Innov. Collab. Keys to Afford. Energy*, vol. 1, pp. 879–893, 2014, doi: 10.2523/iptc-17798-ms.
- [4] S. Güney and A. Atasoy, "Multiclass classification of n-butanol concentrations with k-nearest neighbor algorithm and support vector machine in an electronic nose," *Sensors Actuators, B Chem.*, vol. 166–167, pp. 721–725, 2012, doi: 10.1016/j.snb.2012.03.047.
- [5] M. Yesilbudak, S. Sagioglu, and I. Colak, "A new approach to very short term wind speed prediction using k-nearest neighbor classification," *Energy Convers. Manag.*, vol. 69, pp. 77–86, 2013, doi: 10.1016/j.enconman.2013.01.033.
- [6] B. W. Silverman and M. C. Jones, "E. Fix and J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges (1951)," *Int. Stat. Rev. / Rev. Int. Stat.*, 1989, doi: 10.2307/1403796.
- [7] J. R. Rico-Juan, J. J. Valero-Mas, and J. Calvo-Zaragoza, "Extensions to rank-based prototype selection in k-Nearest Neighbour classification," *Appl. Soft Comput. J.*, vol. 85, p. 105803, 2019, doi: 10.1016/j.asoc.2019.105803.
- [8] "Introduction to k-Nearest-Neighbors - Towards Data Science." [Online]. Available: <https://towardsdatascience.com/introduction-to-k-nearest-neighbors-3b534bb11d26>. [Accessed: 15-Mar-2020].
- [9] F. Report, "SPECIAL CORE ANALYSIS STUDY Heavy Oil Diluent Evaluation," vol. 93308, no. 661, 2011.
- [10] "Aprendizaje Supervisado K - Vecinos más cercanos."
- [11] V. Kuuskraa and M. Wallace, "TECHNOLOGY CO₂-EOR set for growth as new CO₂ supplies emerge," *Oil Gas J.*, vol. 112, no. 5, pp. 92–92, 2014.