



Tipo de artículo: Artículos originales
Temática: Inteligencia Artificial
Recibido: 10/03/2023 | Aceptado: 29/03/2023 | Publicado: 30/03/2023

Identificadores persistentes:
ARK: ark:/42411/s11/a82
PURL: 42411/s11/a82

Predicción de la presión de burbujeo utilizando aprendizaje automático

Prediction of bubble pressure using machine learning

Oscar G. Gil M. ¹[\[0000-0002-4108-2431\]*](https://orcid.org/0000-0002-4108-2431)

¹ Universidad del Zulia (L.U.Z). Maracaibo, Venezuela. profesoroscargil@gmail.com.

Resumen

En el presente estudio se utilizó la colección de algoritmos de aprendizaje automático del programa Weka para predecir la presión de burbujeo de 36 muestras de petróleo, determinando la precisión de sus resultados con el método de prueba validación cruzada de 10 pliegues. Posteriormente, para efectos de comparación, se calcularon las presiones de burbujeo con la correlación generada en el trabajo del cual se tomaron las muestras y sus resultados fueron más precisos que los obtenidos por los algoritmos en 4 de las 7 métricas de rendimiento utilizadas. En virtud de esta situación, y considerando que la correlación fue evaluada con los mismos datos con los que fue generada, se cambió el método de prueba a validación con los datos de entrenamiento y se volvieron a predecir las presiones de burbujeo. En igualdad de condiciones, el aprendizaje automático obtuvo mayor precisión que la correlación en todas las métricas de rendimiento.

Palabras clave: Algoritmos, Aprendizaje automático, Método de prueba, Presión de burbujeo, Weka.

Abstract

In the present study, the collection of machine learning algorithms of the Weka program was used to predict the bubble pressure of 36 oil samples, determining the accuracy of their results with the 10-fold cross-validation test method. Subsequently, for comparison purposes, the bubble pressures were calculated with the correlation generated in the work from which the samples were taken and their results were more precise than those obtained by the algorithms in 4 of the 7 performance metrics used. Due to this situation, and considering that the correlation was evaluated with the same data with which it was generated, the test method was changed to validation with the training data and the bubble pressures were predicted again. Other things being equal, machine learning was more accurate than correlation on all performance metrics.

Keywords: Algorithms, Machine learning, Test method, Bubble pressure, Weka.

Introducción

Las propiedades físicas de los fluidos son de gran importancia en los estudios de ingeniería de petróleo debido a que son necesarias para calcular los hidrocarburos inicialmente en sitio, para simular el comportamiento de los yacimientos y de los pozos, así como también para realizar el diseño de las facilidades de superficie. De todas estas propiedades, la presión de burbujeo (P_b) es probablemente la más importante porque determina la existencia o no de una fase gaseosa que cambia las características del flujo en el yacimiento, en los pozos y en las facilidades de superficie. Adicionalmente, la presión de burbujeo aparece como una discontinuidad, ya que las tendencias con presión de otras propiedades como la relación gas petróleo en solución (R_s) cambian en ese punto [1].

Para determinar la presión de burbujeo y el resto de propiedades se realizan en un laboratorio un conjunto de pruebas comúnmente denominadas análisis PVT, ya que en ellas se analizan las relaciones entre presión, volumen y temperatura de una muestra de los fluidos del yacimiento. Sin embargo, cuando no se tiene esta información experimental (no se puede tomar una muestra que sea representativa, no están garantizados los costos asociados, etc.) se debe recurrir a correlaciones empíricas, ecuaciones de estado o modelos de aprendizaje automático [2]. El uso de modelos de aprendizaje automático en la industria petrolera se ha incrementado substancialmente a raíz de que la cantidad de datos que tienen que manejar, procesar y analizar es cada vez mayor. En este sentido, se deben destacar los esfuerzos dedicados desde finales de los años 1990 para utilizar el aprendizaje automático en la predicción de las propiedades obtenidas de los análisis PVT [3]. Elsharkawy [4] y Gharbi y col. [5, 6] estuvieron entre los primeros que utilizaron modelos de aprendizaje automático para predecir la presión de burbujeo de los fluidos de un yacimiento. Desde entonces, se han presentado muchos estudios que buscan reemplazar a los métodos tradicionales con técnicas de inteligencia artificial/aprendizaje automático debido a su exactitud, confiabilidad, rápida velocidad de respuesta y robusta capacidad de generalización [7] – [10]. En el presente estudio se utilizó la colección de algoritmos de aprendizaje automático del programa Weka [11] para predecir la presión de burbujeo de 36 muestras de petróleo y se determinó la precisión de sus resultados con los métodos de prueba validación cruzada de 10 pliegues y validación con los datos de entrenamiento. Posteriormente, para efectos de comparación, se calcularon las presiones de burbujeo con la correlación generada en el trabajo del cual se tomó la información

de las muestras [12]. Las variaciones que ocasionaron los cambios de método de prueba en los resultados del estudio determinaron la conveniencia de su extensión para incorporar muestras de petróleo de diferentes regiones y composiciones, porque cuando los algoritmos se prueben con datos de diferentes bases químicas se podrá evaluar la capacidad que tuvieron de comprender y aprender patrones durante los entrenamientos [13].

Métodos y Metodología computacional

Construcción de la base de datos.

La base de datos se construyó con información de 36 muestras de petróleo provenientes de 12 yacimientos localizados costa afuera de los Emiratos Árabes Unidos (U.A.E.). Esta selección se produjo en virtud de que la presión de burbujeo es función tanto de la composición del petróleo como de la presión y temperatura del yacimiento, por lo que estos factores pueden ser aproximados utilizando las gravedades específicas del gas y del petróleo (γ_g y γ_o , adimensionales), la relación gas petróleo en solución a presiones mayores o iguales a la de burbujeo (R_{sb} , PCN/BN) y la temperatura del yacimiento (T , °F) [12]. En la **Tabla 1** se presentan algunos parámetros estadísticos de estos datos y en la **Ecuación 1** se describe su relación:

Tabla 1. Parámetros estadísticos de la base de datos.

Parámetro Estadístico	Presión de Burbujeo (Lpcm)	Gravedad Específica del Gas	Gravedad Específica del Petróleo	Relación Gas-Petróleo en Solución (PCN/BN)	Temperatura (°F)
Máximo	4822	1,116	0,9254	3588	306
Mínimo	541	0,746	0,6731	128	190
Promedio	2190,444	0,932	0,831	810,806	241,444
Desv. Estándar	1154,035	0,089	0,037	799,244	24,976
Coef. de Variación	0,527	0,095	0,045	0,986	0,103

$$P_b = f(\gamma_g, \gamma_o, R_{sb}, T) \quad (1)$$

Modelos de aprendizaje automático.

Los modelos de aprendizaje automático desarrollados en este estudio se definen como modelos de regresión, ya que cuando tratan de predecir un caso desconocido producen un resultado numérico (en este caso presión de burbujeo) dentro de un conjunto infinito de posibles resultados [14].

Para desarrollar estos modelos se utilizó la colección de algoritmos del programa Weka (Waikato Environment for Knowledge Analysis), el cual es un software de código abierto emitido bajo la licencia pública general GNU y creado en la universidad de Waikato en Nueva Zelanda. Este contiene herramientas para la preparación, clasificación, regresión, agrupación, minería de reglas de asociación y visualización de datos [11]. Adicionalmente se debe destacar que este programa se considera un punto de referencia en la historia de las investigaciones de minería de datos y aprendizaje automático porque es el único que ha tenido una adopción tan generalizada y se ha mantenido vigente por un período de tiempo tan extenso [15]. Con la base de datos establecida se construyó el archivo Datos.ARFF (Attribute-Relation File Format) utilizado por el programa y sobre el cual se realizaron todas las corridas y sensibilidades. En la **Figura 1** se presenta un ejemplo de las características de uno de los modelos y en el **Anexo 1** se encuentra el archivo.

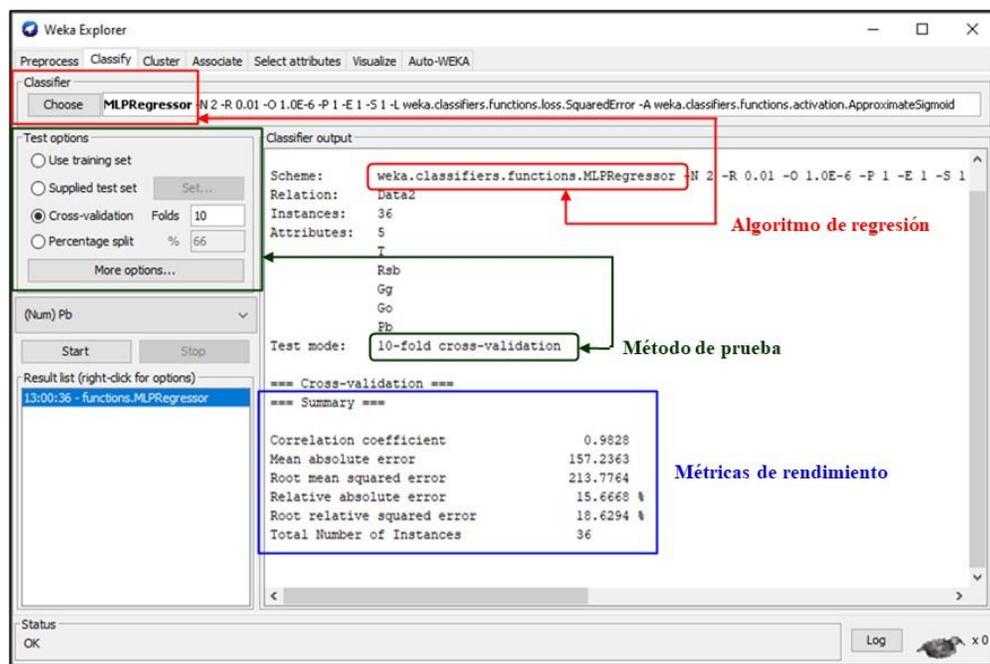


Figura 1. Características de un modelo de regresión de aprendizaje automático con el programa Weka.

Métricas de rendimiento.

El programa Weka utiliza 5 métricas de rendimiento para evaluar la precisión de sus modelos de aprendizaje automático de regresión:

1. Coeficiente de correlación (Correlation coefficient, r^2).
2. Error absoluto medio (Mean absolute error, MAE).
3. Raíz del error cuadrático medio (Root mean squared error, RMSE).
4. Error absoluto relativo (Relative absolute error, RAE).
5. Error absoluto relativo (Relative absolute error, RAE).

En este estudio, además de estas 5 métricas, y en virtud de su uso frecuente en la literatura, también se calcularon los errores porcentuales absolutos (%Eai) de cada predicción y posteriormente, el error porcentual absoluto medio (mean absolute percentage error, MAPE) y la desviación estándar (σ). En el **Anexo 2** se encuentran las ecuaciones utilizadas para los cálculos.

El hecho de utilizar 7 métricas de rendimiento (5 del programa Weka y 2 de la literatura) se debe a que cada una de ellas condensa un gran número de datos en un solo valor, por lo que en realidad ninguna es inherentemente mejor que otra, sino que solamente provee una proyección y enfatiza un aspecto de las características del error del modelo. Por consiguiente, y considerando que diferentes tipos de modelos tienen diferentes distribuciones del error, es evidente que se necesitan diferentes métricas (o incluso una combinación de ellas) para poder evaluar la precisión de los resultados de un modelo de [16, 17, 18].

Resultados y discusión

Aprendizaje automático utilizando validación cruzada de 10 pliegues.

Para determinar la precisión que tuvieron los algoritmos de aprendizaje automático del programa Weka para predecir las 36 presiones de burbujeo de la base de datos se realizó una validación cruzada de 10 pliegues. Los resultados indicaron que los 7 algoritmos de mejor rendimiento fueron: MLPRegressor, AdditiveRegression, MultilayerPerceptron, RBFRegressor, Kstar, RandomizableFilteredClassifier y

M5Rules. Después de esta primera selección, se realizaron sensibilidades en los parámetros internos de estos algoritmos para definir las configuraciones con las que se obtuvieron los mejores resultados. En la **Tabla 2** se presentan las métricas de rendimiento:

Tabla 2. Métricas de rendimiento de los 7 algoritmos de aprendizaje automático de mayor precisión.

Parámetro Estadístico	Validación cruzada 10 pliegues						
	MLPRegressor	AdditiveReg.	MultilayPerc.	RBFRegres.	Kstar	RandomF.	M5Rules
Coefficiente de correlación, r^2	0,9911	0,9864	0,9857	0,9660	0,9527	0,9515	0,9403
Error absoluto medio, MAE	114,6275	146,2712	156,0691	207,7583	252,0619	257,5596	305,2491
Raíz cuadrada del error cuadrático medio, RMSE	156,0982	190,3949	200,2859	295,3871	361,0441	352,6713	407,6970
Error relativo absoluto, RAE	11,5198	14,5742	15,5505	20,7007	25,1151	25,6628	30,4146
Raíz del error cuadrado relativo, RRSE	13,7182	16,5918	17,4538	25,7413	31,4629	30,7333	35,5285
Error absoluto porcentual medio, MAPE	6,2865	8,9878	9,2214	12,8721	15,3572	13,9560	17,3722
Desviación estándar, σ	8,9052	13,4847	13,4929	21,1952	28,1750	21,5082	25,3005

Posteriormente, estas métricas fueron normalizadas y representadas gráficamente de forma adimensional para evaluar sus tendencias de forma comparativa (**Figura 2**).

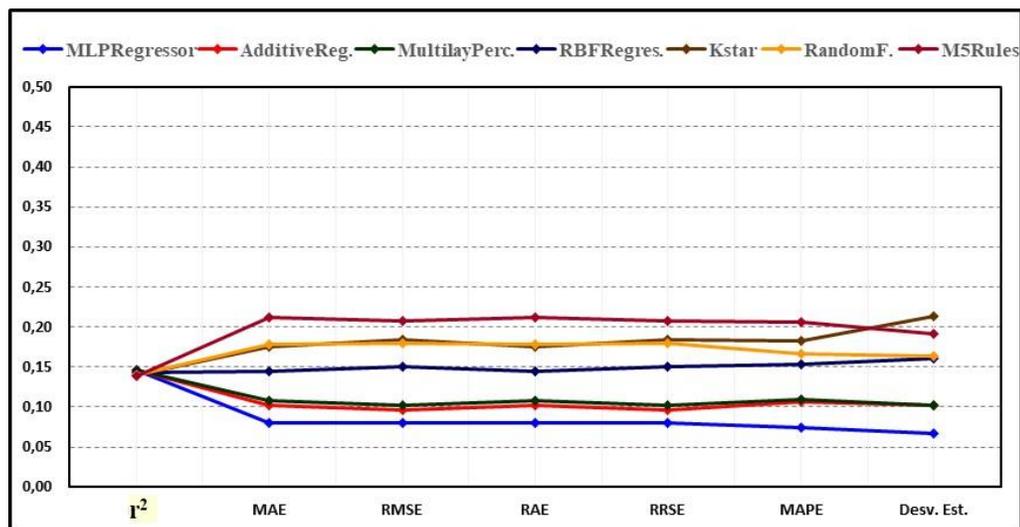


Figura 2. Métricas de rendimiento adimensionales de los 7 algoritmos de aprendizaje automático de mayor precisión.

En la información presentada se observa que con la validación cruzada el algoritmo de mayor precisión fue MLPRegressor, sin embargo, en virtud de que AdditiveRegression y MultilayerPerceptron también obtuvieron buenos resultados y sus tendencias tuvieron un comportamiento similar, se seleccionaron estos 3 para continuar el estudio y se descartó el resto.

Correlación generada utilizando análisis de regresión.

En el trabajo del cual se tomó la información de las 36 muestras de petróleo se realizó un análisis de regresión para generar la correlación que mejor ajustaba las propiedades bajo consideración. En este estudio, para efectos de comparación, se utilizó esa correlación para calcular las presiones de burbujeo y sus resultados fueron más precisos que los obtenidos por los algoritmos de aprendizaje automático en 4 de las 7 métricas de rendimiento: MAE, RAE, MAPE y σ (MLPRegressor lo fue en r^2 , RMSE y RRSE).

Esta elevada precisión que tuvo la correlación se debe tratar con cuidado, porque se consiguió en una prueba con los mismos datos con los que fue generada, los cuales como ya tienen la tendencia regional de los Emiratos Árabes Unidos, “ya conocen” los resultados correctos, pero si la prueba se realiza con muestras de petróleo de diferentes regiones o composiciones (las cuales tienen diferentes bases químicas), se pueden obtener errores significativos [19]. Por el contrario, cuando el aprendizaje automático realiza la validación cruzada de 10 pliegues se está desarrollando un modelo mucho más generalizado, menos sesgado y que no tiene tanta dependencia de los datos utilizados porque se construye 10 veces tomando cada vez 1/10 de los datos para la prueba y el resto para la construcción. Después que el proceso se ha repetido las 10 veces se calcula un promedio con la precisión de cada uno de los modelos [20]. En la **Tabla 3** y **Figura 3** se presenta una comparación de las métricas de rendimiento de los algoritmos de aprendizaje automático y de la correlación:

Tabla 3. Comparación de las métricas de rendimiento de los algoritmos de aprendizaje automático y

Parámetro Estadístico	de la correlación cruzada 10 pliegues			Correlación
	MLPRegressor	AdditiveReg.	MultilayPerc.	
Coefficiente de correlación, r^2	0,9911	0,9864	0,9857	0,9846
Error absoluto medio, MAE	114,6275	146,2712	156,0691	114,4554
Raíz cuadrada del error cuadrático medio, RMSE	156,0982	190,3949	200,2859	203,7676
Error relativo absoluto, RAE	11,5198	14,5742	15,5505	11,5025
Raíz del error cuadrado relativo, RRSE	13,7182	16,5918	17,4538	17,9074
Error absoluto porcentual medio, MAPE	6,2865	8,9878	9,2214	4,5215
Desviación estándar, σ	8,9052	13,4847	13,4929	6,2040

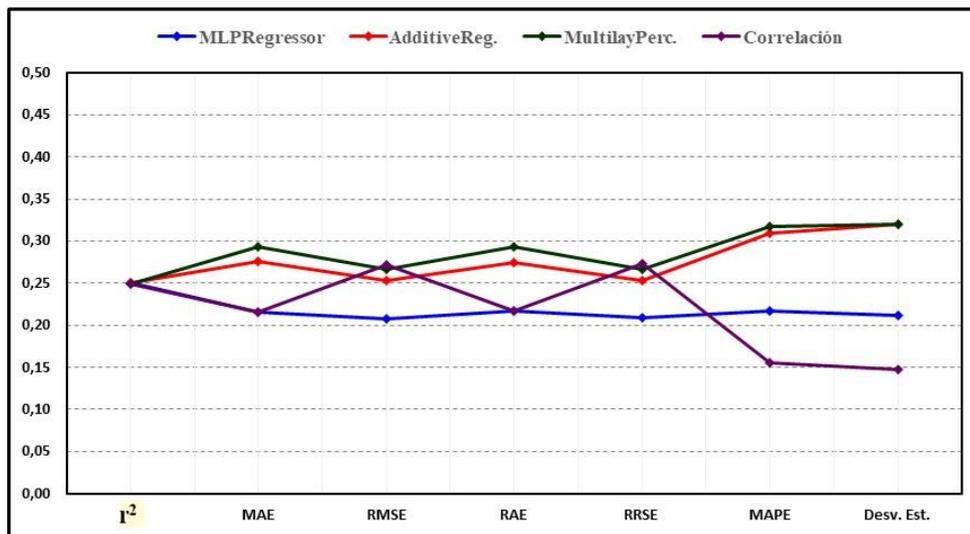


Figura 3. Comparación de las métricas de rendimiento adimensionales de los algoritmos de aprendizaje automático y de la correlación.

Aprendizaje automático utilizando validación con los datos de entrenamiento.

Para comparar en igualdad de condiciones los algoritmos de aprendizaje automático y la correlación se cambió el método de prueba a validación con los datos de entrenamiento y se volvieron a predecir las presiones de burbujeo. De igual manera, también se volvieron a realizar sensibilidades en los parámetros internos de los algoritmos y estas demostraron que existen diferencias entre las configuraciones con las que se obtuvieron

los mejores resultados utilizando validación cruzada de 10 pliegues y las que lo hacen cuando se valida con los datos de entrenamiento. En la **Tabla 4** y **Figura 4** se presenta la comparación de las métricas de rendimiento:

Tabla 4. Comparación de las métricas de rendimiento de los algoritmos de aprendizaje automático y de la correlación.

Parámetro Estadístico	Validación con los datos de entrenamiento			Correlación
	MLPRegressor	AdditiveReg.	MultilayPerc.	
Coefficiente de correlación, r^2	0,9950	0,99997	0,9961	0,9846
Error absoluto medio, MAE	91,5330	5,5375	99,4378	114,4554
Raíz cuadrada del error cuadrático medio, RMSE	114,0809	8,2665	125,2687	203,7676
Error relativo absoluto, RAE	9,1989	0,5565	9,9933	11,5025
Raíz del error cuadrado relativo, RRSE	10,0256	0,7265	11,0088	17,9074
Error absoluto porcentual medio, MAPE	5,4272	0,4119	6,0424	4,5215
Desviación estándar, σ	7,9282	0,7362	8,4451	6,2040

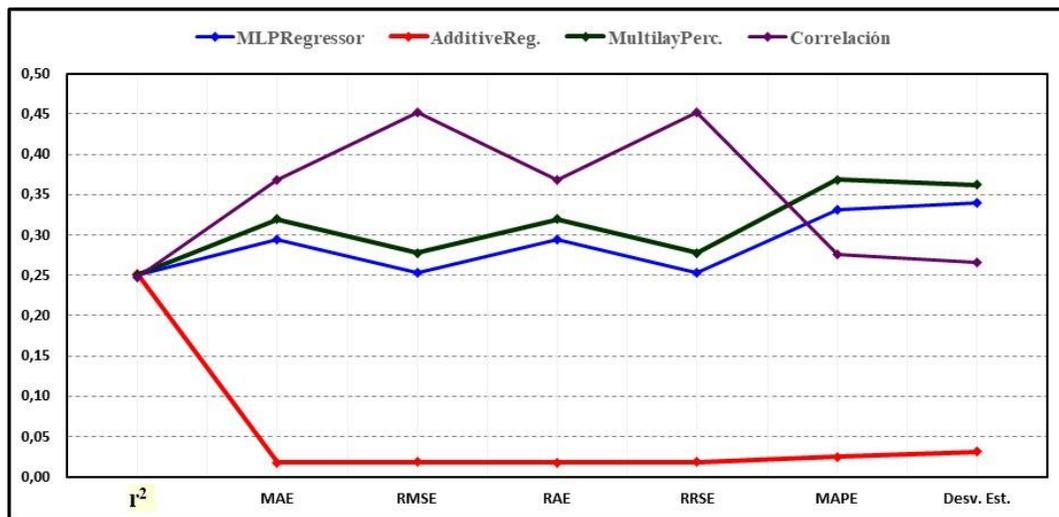


Figura 4. Comparación de las métricas de rendimiento adimensionales de los algoritmos de aprendizaje automático y de la correlación.

En la información presentada se observa que el algoritmo AdditiveRegression obtuvo los mejores resultados en todas las métricas de rendimiento, mientras que MLPRegressor y MultilayerPerceptron fueron más precisos en r^2 , MAE, RMSE, RAE y RRSE que la correlación (esta fue superior en MAPE y σ).

Capacidad de comprender y aprender patrones de los algoritmos de aprendizaje automático.

Los análisis numéricos y gráficos de las métricas de rendimiento de los resultados determinaron que con la validación cruzada el algoritmo MLPRegressor obtuvo los mejores resultados en 3 métricas de rendimiento (la correlación lo hizo en 4), y que posteriormente, cuando se cambió el método de prueba a validación con los datos de entrenamiento fue el algoritmo AdditiveRegression el que obtuvo mejores resultados en todas las 7 métricas de rendimiento (MLPRegressor y MultilayerPerceptron fueron superiores a la correlación en 5). Las variaciones que ocasionaron los cambios de método de prueba en los resultados del estudio determinan la conveniencia de su extensión para incorporar muestras de petróleo de diferentes regiones y composiciones, porque cuando los algoritmos se prueben con datos de diferentes bases químicas se podrá evaluar la capacidad que tuvieron de comprender y aprender patrones durante los entrenamientos.

Conclusiones

Los algoritmos de aprendizaje automático del programa Weka que obtuvieron mayor precisión para predecir la presión de burbujeo de las 36 muestras de petróleo utilizando validación cruzada de 10 pliegues fueron MLPRegressor, AdditiveRegression y MultilayerPerceptron. Cuando se utilizó validación cruzada de 10 pliegues el algoritmo MLPRegressor obtuvo resultados más precisos en 3 métricas de rendimiento y la correlación generada con análisis de regresión lo hizo en 4. Cuando se utilizó validación con los datos de entrenamiento el algoritmo AdditiveRegression obtuvo resultados más precisos en todas las métricas de rendimiento. Las sensibilidades realizadas en los parámetros internos de los algoritmos de aprendizaje automático demostraron que existen diferencias entre las configuraciones con las que se obtuvieron los mejores resultados utilizando validación cruzada de 10 pliegues y las que lo hacen cuando se valida con los datos de entrenamiento. Los resultados del estudio determinaron la conveniencia de su extensión para

incorporar muestras de petróleo de diferentes regiones y composiciones, porque cuando los algoritmos se prueben con datos de diferentes bases químicas se podrá evaluar la capacidad que tuvieron de comprender y aprender patrones durante los entrenamientos.

Referencias

- [1] X. Yang, B. Dindoruk and L. Lu, “A comparative analysis of bubble point pressure prediction using advanced machine learning algorithms and classical correlations”, *Journal of Petroleum Science and Engineering*, vol. 185, 106598, 2020. Available: <https://doi.org/10.1016/j.petrol.2019.106598>.
- [2] M. M. Almashan, Z. Arsalan, Y. Narusue and H. Morikawa, “Estimating Pressure-Volume-Temperature Properties of Crude Oil Systems Using Boosted Decision Tree Regression”, *Journal of the Japan Petroleum Institute*, vol. 65, n°. 6, pp. 221-232, 2022. Available: <https://doi.org/10.1627/jpi.65.221>.
- [3] K. Ghorayeb, A. Mawlod, A. Maarouf, Q. Sami, N. El Droubi, R. Merrill, O. El Jundi and H. Mustapha, “Chain-based machine learning for full PVT data prediction”, *Journal of Petroleum Science and Engineering*, vol. 208, Part D, 109658, 2022. Available: <https://doi.org/10.1016/j.petrol.2021.109658>.
- [4] A. M. Elsharkawy, "Modeling the Properties of Crude Oil and Gas Systems Using RBF Network." Paper presented at the SPE Asia Pacific Oil and Gas Conference and Exhibition, Perth, Australia, October 1998. Available: <https://doi.org/10.2118/49961-MS>.
- [5] R. B. Gharbi and A. M. Elsharkawy, “Neural network model for estimating the PVT properties of Middle East crude oils”. Paper presented at the Middle East Oil Show and Conference, Bahrain, March 1997. Available: <https://doi.org/10.2118/37695-MS>.
- [6] R. B. Gharbi, A. M. Elsharkawy and M. Karkoub, “Universal Neural-Network-Based Model for Estimating the PVT Properties of Crude Oil Systems”. *Energy Fuels*, vol. 13, pp. 454–458, 1999. Available: <https://doi.org/10.1021/ef980143v>.

- [7] S. Alatefi and A. M. Almeshal. “A New Model for Estimation of Bubble Point Pressure Using a Bayesian Optimized Least Square Gradient Boosting Ensemble”. *Energies*, vol. 14, n°. 9, pp. 2653, 2021. Available: <https://doi.org/10.3390/en14092653>.
- [8] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist and H. Oza. “Application of machine learning and artificial intelligence in oil and gas industry”. *Petroleum Research*, vol. 6, n°. 4, pp. 379-391, 2021. Available: <https://doi.org/10.1016/j.ptlrs.2021.05.009>.
- [9] M. Ahmadi, M. Pournik and S. Shadizadeh. “Toward connectionist model for predicting bubble point pressure of crude oils: Application of artificial intelligence”. *Petroleum*, vol. 1, n°. 4, pp. 307-317, 2015. Available: <https://doi.org/10.1016/j.petlm.2015.08.003>.
- [10] F. Alakbari, M. Mohyaldinn, M. Ayoub, A. Muhsan and I. Hussein. “A reservoir bubble point pressure prediction model using the Adaptive Neuro-Fuzzy Inference System (ANFIS) technique with trend analysis. *PLoS ONE* vol. 17, n°. 8, e0272790, 2022. Available: <https://doi.org/10.1371/journal.pone.0272790>.
- [11] Weka 3-Data Mining with Open Source Machine Learning Software in Java. Available online: <https://www.cs.waikato.ac.nz/ml/weka/>.
- [12] R. A. Al-Mehaideb, “Improved PVT Correlations for UAE Offshore Crudes”, *Journal of The Japan Petroleum Institute*, vol. 40, n°. 3, pp. 232-235, 1997. Available: <https://doi.org/10.1627/jpi1958.40.232>.
- [13] S.M. Macary and M.H. El-Batanoney “Derivation of PVT Correlations for the Gulf of Suez Crude Oils”, *Journal of The Japan Petroleum Institute*, vol. 36, n°. 6, pp. 472-478, 1993. Available: <https://doi.org/10.1627/jpi1958.36.472>.
- [14] IArtificial.net: ¿Clasificación o Regresión? [Online]. Available: <https://www.iartificial.net/clasificacion-o-regresion/>.

- [15] M. Senthamilselvi and P.S.S. Akilashri, "A Comparative Study on Weka, orange Tool for Mushroom data Set", International Journal of Computer Sciences and Engineering, vol. 06, n°. 11, pp. 231-236, 2018. Available: www.ijcseonline.org.
- [16] T. Chai and R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literatura", Geoscientific Model Development, vol. 7, n°. 3, pp. 1247–1250, 2014. Available: <https://doi.org/10.5194/gmd-7-1247-2014>, 2014.
- [17] T. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not", Geoscientific Model Development, vol. 15, n°. 14, pp. 5481–5487, 2022. Available: <https://doi.org/10.5194/gmd-15-5481-2022>.
- [18] S. Taipe y G. Ampuño, "Modelo del proceso de producción de energía en centrales de generación térmica considerando el perfil de funcionamiento", Ciencia Latina Revista Científica Multidisciplinar, vol. 6, n°. 4, pp. 5541-5560, 2022. Available: https://doi.org/10.37811/cl_rcm.v6i4.3032.
- [19] B. Moradi, E. Malekzadeh, M. Amani, F. Boukadi, and R. Kharrat. "Bubble Point Pressure Empirical Correlation." Paper presented at the Trinidad and Tobago Energy Resources Conference, Port of Spain, Trinidad, June 2010. Available: <https://doi.org/10.2118/132756-MS>.
- [20] Data science: Validación cruzada K-Fold [Online]. Available: <https://datascience.eu/es/aprendizaje-automatiko/validacion-cruzada-de-k-fold/>.

Anexos

Anexo 1. Archivo Datos.ARFF utilizado por el programa Weka.

```
% Título
@relation Datos

% Lista de atributos y su tipo
@attribute T numeric
@attribute Rsb numeric
@attribute Gg numeric
@attribute Go numeric
@attribute Pb numeric

% T, Rsb, Gg, Go, Pb
@data
250,603,0.923,0.8519,2425
234,265,1.0112,0.855,1190
275,563,0.9633,0.8076,1787
220,349,1.0304,0.827,1197
255,1985,0.9146,0.8265,3796
244,476,1.0165,0.8348,1625
234,259,0.9954,0.8483,1062
230,226,0.9859,0.8676,1030
268,1146,0.8358,0.8114,3399
230,252,1.0587,0.8597,994
234,589,0.8969,0.8328,2061
239,448,1.0023,0.8418,1591
226,861,0.8504,0.8633,3184
250,965,0.8382,0.8198,3202
240,1057,0.9159,0.8217,2946
230,903,0.833,0.8241,2944
220,264,0.9591,0.8319,1179
252,1894,0.8029,0.8035,4627
239,371,0.9563,0.866,1490
306,3588,0.9646,0.7857,3402
300,3348,0.9663,0.7866,3406
232,265,0.9944,0.8524,1104
280,1243,0.7969,0.6731,2703
230,993,0.8544,0.8179,3172
240,1214,0.746,0.8444,4822
266,372,1.0817,0.9254,1430
250,399,0.8541,0.8263,1920
275,423,1.116,0.812,1360
254,277,0.8675,0.8435,1345
205,128,0.9691,0.8155,541
220,727,0.8655,0.8184,2509
219,1240,0.8164,0.8363,4004
220,141,1.0461,0.8433,590
220,730,0.8993,0.827,2417
215,294,0.9878,0.867,1261
190,331,0.9499,0.8338,1141
```

Anexo 2. Ecuaciones utilizadas para calcular las 7 métricas de rendimiento y los errores porcentuales absolutos.

Nomenclaturas:

X_i = Valores reales de la presión de burbujeo.

\bar{X} = Promedio aritmético de los valores reales de la presión de burbujeo.

Y_i = Valores calculados de la presión de burbujeo.

\bar{Y} = Promedio aritmético de los valores calculados de la presión de burbujeo.

n = Número de muestras.

A.2.1. Coeficiente de correlación (Correlation coefficient, r^2).

$$r^2 = \frac{\sum_{i=1}^n (X_i - \bar{X}) * (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

A.2.2. Error absoluto medio (Mean absolute error, MAE).

$$MAE = \frac{1}{n} \sum_{i=1}^n |X_i - Y_i|$$

A.2.3. Raíz del error cuadrático medio (Root mean squared error, RMSE).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2}$$

A.2.4. Error absoluto relativo (Relative absolute error, RAE).

$$RAE = \frac{\sum_{i=1}^n |X_i - Y_i|}{\sum_{i=1}^n |X_i - \bar{X}|}$$

A.2.5. Raíz del error cuadrado relativo (Root relative squared error, RRSE).

$$RRSE = \sqrt{\frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

Errores porcentuales absolutos (%Eai).

$$\%Eai = \left| \frac{X_i - Y_i}{X_i} \right| * 100$$

A.2.6. Error porcentual absoluto medio (mean absolute percentage error, MAPE).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|X_i - Y_i|}{X_i} * 100$$

A.2.7. Desviación estándar (σ).

$$\sigma = \sqrt{\frac{1}{n-1} \left[\left(\frac{X_i - Y_i}{X_i} \right) * 100 \right]^2}$$

Roles de Autoría

Oscar Gil: Conceptualización, Curación de datos, Análisis formal, Investigación, Metodología, Software, Validación, Redacción - borrador original.