Machine Learning Applied to Viscosity Prediction: A Case Study

GIL VERA VICTOR DANIEL Department of Engineering, University Catholic Luis Amigó, COLOMBIA

Abstract: - Viscosity emerges as a physical property of primary importance in the modeling of flow within a porous medium, as well as in the processes of production, transport, and refining of crude oils. The direct measurement of viscosity is carried out through laboratory tests applied to samples extracted from the bed of a well, being these samples characterized by their difficult collection and the considerable time lapse required for their acquisition. Several techniques have been developed to estimate viscosity, among which the empirical correlation with Nuclear Magnetic Resonance logs stands out. This study presents a methodology for creating a representative predictive viscosity model, adapted to specific reservoir conditions, using measurements and well logs using machine learning techniques, in particular, Support Vector Machines (SVM). It is concluded that SVM trained with a polynomial kernel ($R^2 = 0.947$, MSE = 631.21, MAE = 15.16) exhibits superior performance compared to SVM trained with linear and RBF kernels. These results suggest that SVMs constitute a robust machine-learning technique for predicting crude viscosity in this context.

Key-Words: - Forecasting, Gravity, Machine Learning, Nuclear Magnetic Resonance, Oil, Permeability, Porosity, Regression, Saturation, Viscosity.

Received: July 23, 2023. Revised: October 18, 2023. Accepted: November 19, 2023. Published: December 31, 2023.

1 Introduction

Viscosity, stipulated as in literature, [1], characterizes the fluid's opposition to shear stress or flow and holds paramount significance in the computational representation of engineering procedures spanning the entirety of the petroleum encompassing fluid processes from sector, extraction to the refinement stage. Accurately estimating the propelling forces driving fluid flow necessitates the accessibility of viscosity data contingent on pressure, temperature, and density. As a result, hydraulic computations pertinent to fluid production and conveyance systems, as well as flow simulations within porous media, hinge upon the capability to prognosticate fluid viscosity under defined procedural circumstances, [2].

The characteristic under consideration assumes a pivotal role in formulating and advancing procedures aimed at recuperating, enhancing, and refining viscous crude oils. Owing to their elevated viscosity, these oils encounter notable impediments in spontaneous migration towards the wellbore, thereby rendering conventional production techniques insufficient for their extraction, [3]. By contrast, normal oils have viscosities of around 1 (cP) to 10 (cP), whereas heavy crudes can exceed the 1 million (cP) limit in normal circumstances. Since these crudes make up as much as 70% of the world's petroleum reserves, specific recovery techniques have been created for the reservoirs that contain these oils, which makes viscosity mitigation tactics necessary, [4].

Viscosity measurements are directly obtained by the study of crude oil samples that are extracted from the wellbore, [5]. These samples must be of a significant caliber to provide timely, accurate, and useful results for effective manufacturing systems. Fluid characteristics are altered as a result of the large temperature and pressure changes that the reservoir fluids undergo during the collection process. These differences from in situ conditions are substantial. The samples are forwarded to labs for analysis once they are collected. Nevertheless, this procedure may cause a delay in data availability, which might impair the capacity to decide on development plans quickly, [6].

Some other approaches to the assessment of viscosity include petro-physical Nuclear Magnetic Resonance (NMR) logging. This method is used for the thorough assessment of characteristics including saturation, porosity, and permeability as well as the characterization of various fluids in geological formations, regardless of lithology, [7]. The underlying physical process is the induction of a magnetic field that drives the fluids in the porous medium's magnetic cores into action. These cores

interact, especially with hydrogen cores, to collect energy and then release it again.

Relaxation periods are defined as the rate at which the magnetic signals connected to this energy's re-emission diminish exponentially with time, [8]. Measurements of oil viscosity and this degradation pattern have been found to correlate empirically, [9]. The link between temperature and hydrogen index is dependent on several factors. Unfortunately, these relationships do not work well enough to forecast heavy crude viscosities, [10].

Complex and nonlinear engineering problems have been solved in the scientific literature by using Machine Learning techniques including Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Linear Regression. These methods have demonstrated impressive effectiveness in the modeling of issues with many variables, nonlinearity, and large amounts of data, [11].

machine In this study. learning-more particularly, Support Vector Machines-is used to propose a predictive viscosity model based on Nuclear Magnetic Resonance logs (SVM). A database composed of 366 logs was employed, using API gravity, gas-liquid ratio, sampling depth, temperature, pressure, and X and Y geographic coordinates as predictor variables. The target The variable was viscosity (cP). Python programming language was employed alongside the application of the Support Vector Machines (SVM) Machine Learning methodology.

The subsequent sections of the manuscript are organized as follows: Section 2 delineates the problem formulation, Section 3 expounds upon the theoretical framework, Section 4 outlines the methodology, and Section 5 details the presentation and discussion of results. The paper culminates with a conclusion.

2 **Problem Formulation**

Viscosity, understood as the resistance of a fluid to shear stress, [12], is used for the modeling of engineering processes present in all aspects of the petroleum industry, [13], from production to the refining of fluids, [14]. Viscosity values at the given pressure, temperature, and density are required to estimate the driving forces for the fluid flow, [15]. Therefore, hydraulic calculations for production and transport systems, as well as the modeling of flow in porous flow modeling in porous media depend on the prediction of the fluid viscosity at the process conditions, [16].

One of the methods to estimate viscosity is the Nuclear Magnetic Resonance (NMR), [17], which is

used to estimate properties such as porosity, permeability, saturation, and characterization of the different fluids present in the geological formation independent of lithology, [18]. The physical principle that governs it consists of an induced magnetic field that stimulates the magnetic nuclei of the fluids housed in the porous medium, [19], which absorb and re-emit energy through interaction with other nuclei of the fluid components, specifically hydrogen, [20].

In recent years, complex and non-linear engineering problems have been solved with Machine Learning techniques such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Linear Regression, as they have shown satisfactory performance in modeling satisfactory performance in the modeling of problems with multiple variables, non-linear and with large volumes of information, [21]. For this reason, the objective of the present research is to develop a methodology to develop a predictive model of viscosity from petro-physical logs and viscosity measurements, by implementing the Machine Learning workflow.

3 Background

-Reservoir Fluids - Petroleum: petroleum constitutes a sophisticated amalgamation of diverse encompassing constituents. various natural hydrocarbon compounds. organic compounds nitrogen, containing oxygen, and sulfur. nonhydrocarbons, and trace quantities of metallic elements such as nickel, iron, and vanadium, [22]. The composition and characteristics of these fluids exhibit considerable heterogeneity contingent upon geological formations, including factors such as density, viscosity, and volatility.

-Presence in the Reservoir: petroleum manifests itself within the reservoir as either liquid oil or natural gas. As crude oil pressure diminishes, light hydrocarbons and nonhydrocarbons separate from the liquid oil reservoir fluids, transitioning into a gaseous phase, [23]. Essentially, natural gas mixtures comprise light alkanes (ranging from methane to n-butane) and nonhydrocarbons, such as nitrogen (N₂), carbon dioxide (CO₂), hydrogen sulfide (H₂S), helium (He), and trace amounts of water vapor. Moreover, they may contain minimal quantities of heavier hydrocarbon components, other gaseous non-hydrocarbons, and inert gases, [24].

-Importance of Physical Properties in Process Modeling: the physical properties of petroleum fluids, irrespective of phase, play a pivotal role in process modeling. Therefore, it is imperative to consider diverse classifications of crude oil, [25]. Notably, a primary classification is based on fluid volatility, which correlates with specific gravity (° API) and the quantity of dissolved gas (GOR) under reservoir conditions. In this context, five types of reservoir fluids are delineated: black oil, volatile oil, gas condensate, wet gas, and dry gas, [13].

These crude oils can further be categorized based on their specific gravity (Table 1) or by their density and viscosity (Table 2).

ruble 1. Crude on Typology				
Type of crude oil	Density (Kg/m ³)	°API		
Light	< 870	> 31.1		
Medium	870 - 920	22.3 - 31.1		
Heavy	920 - 1000	10.0 - 22.3		
Extra-heavy	> 1000	< 10.0		

Table 1. Crude Oil Typology

Table 2. Heavy Crude Oil Classification

Tuble 2: Heavy Crude On Classification					
Type of crude	Viscosity	Density	°API		
oil	(cP)	(Kg/m ³)			
Light	< 100	< 934	> 20		
Heavy	100 - 100000	934 - 1000	10-20		
Bitumen	>100000	>1000	< 10		

Crude oil typologies are also segregated according to their extraction method. In this instance, conventional oils denote light and medium category oils, characterized by relatively low viscosities, obtained by traditional recovery methods. In contrast, unconventional crudes comprise high-viscosity oils, such as heavy, extraheavy, and bituminous oils, or light oils hosted in very low permeability rock formations, [26].

The hydrocarbon sampling procedure plays a crucial role in reservoir development decisionmaking. Two main approaches for acquiring such samples stand out: downhole sampling and surface sampling. The former involves the introduction of sampling tools through a production test string (DST), wireline, or tubing to the productive region. In situations where the well has not been cased or the hole remains open, sampling can be carried out using the modular formation dynamics tester (MDT), [27].

Cased hole sampling incorporates the Cased Hole Dynamics Tester (CHDT), which seals a pack against the borehole or casing wall and then presses a probe against the formation. When pumping is initiated, the fluid contained in the rock is drawn out through the probe's intake port. This type of sampling allows, in general, the preservation of the sample in conditions as close as possible to those of the reservoir, [28]. On the other hand, surface sampling is most often carried out at the separator under stable flow conditions. It involves the collection of gas and liquid samples and can be performed throughout the productive life of the well, [29].

After sample collection, samples are subjected to a series of laboratory tests for fluid characterization. Standardized analyses in this context include composition, density, gas-oil ratio, saturation pressure, asphaltene stability, and viscosity, [30]. Viscosity is the resistance of a fluid to shear stress. In Newton's viscosity law, it is defined in terms of the velocity gradient (u) and the shear stress (τxy) as follows:

$$\tau_{xy} = -\mu \, \frac{\partial \mu_x}{\partial_y} \tag{1}$$

The generation of momentum transfer is attributable to shear stress, and viscosity is defined as the proportionality constant between the driving force and the subsequent velocity gradient, [31]. Crude viscosity is evaluated using various laboratory devices, such as rheometers and viscometers designed to operate at high pressures. In addition, the use of other instruments such as hydrometers, pressure pumps, and temperature baths, among others, is necessary, [32].

4 Methodology

The dataset consisted of 7 attributes and 366 logs. It contains, geographic coordinates and depth data, physicochemical characteristics of the crude oil, and viscosity tests with its corresponding pressure and temperature. The dataset used is available in, [33]. Table 3 presents the complete description of the same.

The CRISP-DM methodology was used to build the model. Figure 1 shows all the phases that comprise it, each of them is described below:

- *Stage 1 Business Understanding:* In this phase, a comprehensive understanding of the business objectives is pursued. Critical factors related to the desired results are identified, the project objectives are established, and a plan is drawn up that defines the steps to be followed, the tools and techniques required, and the success criteria that will determine the achievement or failure of the proposed objectives, [34]. Likewise, success criteria are defined, which will determine the achievement or failure of the proposed objectives, [35].

- Stage 2 Data Understanding: Data collection, identification of quality problems in the data, and obtaining the first relevant knowledge are carried

out. During this phase, subsets of data of interest for the formulation of new hypotheses may be identified, [36].

- *Stage 3 Data Preparation:* The raw data are cleaned and converted before the processing and analysis stage. The final objective of this phase is to obtain the final data on which the models will be applied; the universe of data to work with is established and debugged, [37].

- *Stage 4 Modeling:* With the data normalized and cleaned, we proceed to the construction of the models with the optimal parameters (cross-validation), [38].

- *Stage 5 Evaluation:* The performance of the models built from the performance metrics is evaluated and the optimal one is selected, [39].

- *Stage 6 Implementation:* The optimal model identified in the previous phase is implemented, either through a graphical user interface or directly from software (Python, R, Matlab, etc.), [40].

1					
Variable	Description				
API Gravity (V1)	Crude oil density indicator at standard conditions. Samples within the range of 9.9 to 13.8 °API have been incorporated, covering crudes from extra-heavy to heavy category.				
Gas Oil Ratio (V2)	The ratio of the volume of gas released by the fluid to the volume of oil at standard conditions (scf/bbl). The data set contains information from tests on dead crude and live samples in the range of 5.3 to 24.8 scf/bbl.				
Sampling Deth TVDES (V3)	Vertical depth to sea level of the well in feet (ft). Samples were obtained in the depth range of -5329 to -6462 ft.				
Viscosity Test Temperature (V4)	Temperature in (°F) at which the viscosity measurement was performed. Tests were performed from 140 to 350 °F.				
Viscosity Test Pressure (V5)	Pressure (psi) at which the viscosity measurement was made. The evaluation range is 15 to 4015 psi.				
Geographic coordenate X (V6)	east-west geographic coordinate of location of the well where the crude il sample was obtained.				
Geographic coordenate Y (V7)	Geographic Y north-south coordinates of the location of the well where the oil sample was obtained.				
Viscosity (V8)	Dynamic viscosity (cP) is measured at constant temperature and variable pressure.				

Table 3. Description

Note: All variables are continuous numerical.



Fig. 1: CRISP-DM Methodology

In Machine Learning, the quality and volume of data influence the accuracy of predictive models. To guarantee the model's applicability, it is also crucial to validate the model and understand the findings. Iterations and tweaks may be necessary in this procedure based on the validation findings. Since this study involves regression, the machine learning (ML) approach of support vector machines with several kernel types (linear, polynomial, and RBF radial basis functions) was applied. Equation (2) presents the general equation for Support Vector Machines (SVM).

Equations (3), (4), and (5) illustrate the various kernel types: polynomial, linear, and RBF. The kernel function is represented by K, the intercept by b, the dual coefficients by Alphas, and the optimization parameters by C, γ , r, and d.

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x) + b$$
 (2)

Lineal Kernel

$$K(x_n, x_i) = (x_n, x_i) \tag{3}$$

RBF Kernel

$$K(x_n, x_i) = exp(-\gamma ||x_n - x_i||^2) + C$$
 (4)

Polynomial Kernel

$$K(x_{n}, x_{i}) = (\gamma (x_{n}, x_{i}) + r)^{d}$$
(5)

5 Results and Discussion

Table 4 depicts the correlation matrix, illustrating the interrelationships among eight variables denoted as V1 to V8. The correlation coefficients within the range of -1 to 1 signify the strength and direction of

associations, with a value of 1 denoting a flawless positive correlation, -1 indicating a flawless negative correlation, and 0 signifying the absence of correlation.

Table 4. Correlation matrix

	V1	V2	V3	V4	V5	V6	V7	V8
V1	1,00							
V2	-0,03	1,00						
V3	0,04	-0,08	1,00					
V4	-0,03	-0,05	0,03	1,00				
V5	-0,03	0,09	0,09	0,05	1,00			
V6	0,03	-0,02	-0,03	0,14	0,00	1,00		
V7	0,06	-0,02	0,04	0,13	0,16	0,12	1,00	
V8	-0,04	-0,08	-0,02	-0,02	-0,02	0,04	-0,03	1,00
Note	V1 -	APL C	ravit	V^2	- Gas	Oil R	atio	V_{2} –

Note: V1 = API Gravity, V2 = Gas Oil Ratio, V3 = Samplig Deth TVDES, V4 = Test Temperature, V5 = Test Pressure, V6 = X, V7 = Y, V8 = Viscosity.

These correlations provide insights into the linear relationships between pairs of variables. Keep in mind that correlation does not imply causation and other statistical methods may be needed for a more comprehensive analysis.

Table 5 presents the results of the different trained models, Accuracy, Recall, precision, and F1-Score metrics were calculated. The fit tests were R², MSE, and MAE. The polynomial kernel performed better than the other two (higher R² and lower MSE and MAE).

Table 5. Results				
	SVM	SVM	SVM	
Metrics	Kernel	Kernel	Kernel	
	Lineal	RBF	Polynomial	
R ²	0.879	0.826	0.947	
MSE	1450.51	2088.98	631.21	
MAE	25.82	24.15	15.16	

The Support Vector Machine (SVM) model with a polynomial kernel has shown a remarkably superior performance in terms of fit metrics compared to other models evaluated. This finding implies that the complexity and nonlinear interactions seen in the examined data have been well captured by the kernel selection.

6 Conclusion

There is a clear relationship of viscosity concerning temperature, viscosity decreases critically with increasing temperature, but concerning depth viscosity increases. This is because the change in viscosity at reservoir conditions responds more to a change in fluid density than to in-situ temperature.

Viscosity information is dependent on many other characteristics apart from geographical position and depth. For future research, it is suggested that more information is suggested to correlate viscosity behavior with more chemically influential characteristics.

References:

- Bryan, J., D. Moon, and A. Kantzas, In situ viscosity of oil sands using low field NMR, *Journal of Canadian Petroleum Technology*, 2005. Vol.44, No.9.
- [2] Freedman, R. and Heaton, N. Fluid characterization using nuclear magnetic resonance logging. *Petrophysics*, 2004. Vol.45, No. 3.
- [3] Alazard, N. and Montadert, L. Oil resources for the next century: What's ahead? *Nonrenewable Resources*, Vol.2, No.3, 1993, pp. 197-206.
- [4] Betancourt, S., Davies, T., Kennedy, R., Dong, C., Elshahawi, H., Mullins, O.C., Nighswander, J. and O'Keefe, M. Advances in the measurement of fluid properties, *Spanish Oilfield Review*. Schlumberger, 2007.
- [5] Coates, G.R., Xiao, L. and Prammer, M.G. NMR logging: principles and applications. Vol. 234, 1999: *Haliburton Energy Services Houston*.
- [6] Dunn, K.J., Bergman, D. J. and LaTorraca, G. A. Nuclear magnetic resonance: Petrophysical and logging applications. 2002: *Elsevier*.
- [7] Morriss, C. E., Freedman, R., Straley, C., Johnston, M., Vinegar, H. J. and Tutunjian, P.N. Hydrocarbon saturation and viscosity estimation from NMR logging in the Belridge Diatomite. in SPWLA 35th Annual Logging Symposium. 1994. Society of Petrophysicists and Well-Log Analysts.
- [8] LaTorraca, G. A., Dunn, K. J., Webber, P. R., Carison, R. M. and Stonard, S. W. Heavy oil viscosity determination using NMR logs. *In* SPWLA 40th Annual Logging Symposium. 1999. Society of Petrophysicists and Well-Log Analysts.
- [9] Yang, Z., Viscosity Evaluation of Heavy Oils from NMR Well Logging. 2011.10. Jang, J.S., ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Management and Cybernetics*, Vol.23, No.3, 1993, pp. 665-685.
- [10] Speight, J.G., The chemistry and technology of petroleum. 2014: *CRC Press*.
- [11] Kidnay, A.J., Parrish, W. R. and McCartney, D. G. Fundamentals of natural gas processing. 2011: CRC Press.

- [12] McCain Jr, W., The Properties of Petroleum Fluids, second edition. Tulsa, *Oklahoma: PennWell Publishing Company*, 1990.
- [13] Okandan, E., Heavy crude oil recovery. Vol. 76, 2012: Springer Science & Business Media.
- [14] Hein, F.J., Heavy oil and oil sands in North America: an overview & summary of contributions. *Natural Resources Research*, Vol.15, No.2, 2006, pp. 67-84.
- [15] Zéberg-Mikkelsen, C.K., Quiñones-Cisneros, S.E. and Stenby, E. H. Viscosity prediction of hydrocarbon mixtures based on the friction theory. *Petroleum Science and Technology*, Vol.19, No 8, 2001, pp. 899-909.
- [16] Nicot, B., Fleury, M. and Leblond, J. A New Methodology for Better Viscosity Prediction Using Nmr Relaxation. In SPWLA 47th Annual Logging Symposium. 2006. Society of Petrophysicists and Well-Log Analysts.
- [17] Cheng, Y., Kharrat, A. M., Badry, R. and Kleinberg, R. L. Power-law Relationship between the Viscosity of Heavy Oils and NMR Relaxation. *In SPWLA 50th Annual Logging Symposium*. 2009. Society of Petrophysicists and Well-Log Analysts.
- [18] Nikravesh, M., Soft computing-based computational intelligence for reservoir characterization. *Expert Systems with Applications*, vol. 26, No.1, 2004, pp. 19-38.
- [19] Smola, A. J. and Schölkopf, B. A tutorial on support vector regression, *Statistics and computing*, Vol.14, No. 3, 2004, pp. 199-222.
- [20] Alboudwarej, H, Felix, J.J, Taylor, S. The importance of heavy oil, *Oilfield Review*, Vol.18, 2006, pp. 38–59.
- [21] Dasgupta, T. and Mukherjee, S. Sediment compaction and applications in petroleum geoscience. 2020: Springer.
- [22] Schlumberger, Oilfield Glossary, 2020, [Online]. <u>https://glossary.slb.com/</u> (Accessed Date: September 19, 2023).
- [23] Shukla, A.K. and Wiley, J. Analytical Characterization Methods for Crude Oil and Related Products. 2018: *Wiley Online Library*.
- [24] Bloembergen, N., Purcell, E. M. and Pound, R. V. Relaxation effects in nuclear magnetic resonance absorption. *Physical Review*, Vol.73, No.7, 1948, pp. 679.
- [25] Freedman, R., Formation evaluation using magnetic resonance logging measurements, Google Patents, 2001, [Online]. <u>https://patents.google.com/patent/US6229308</u> <u>B1/en</u> (Accessed Date: September 12, 2023).
- [26] Bryan, J., Kantzas, A. and Bellehumeur, C. Oil-viscosity predictions from lowfield NMR

measurements. SPE Reservoir Evaluation & Engineering, Vol.8, No. 1, 2005, pp. 44-52.

- [27] Straley, C., Rossini, D., Vinegar, H., Tutunjian, P. and Morriss, C. Core analysis by low-field NMR. *The log analyst*, Vol.38, No.2, 1997
- [28] Bird, R., Stewart, W. and Lightfoot, E. Transport Phenomena, John Wiley & Sons, New York, NY, USA. 2002.
- [29] Kenyon, W., Petrophysical principles of applications of NMR logging. *The Log Analyst*, Vol.38, No. 2, 1997
- [30] Brown, R., Proton relaxation in crude oils. *Nature*, 1961. Vol.189, No.4762, p. 387-388.
- [31] Vinegar, H. NMR fluid properties: NMR Short Course. *In SPWLA 36th Annual Symposium*. 1995.
- [32] Zhang, Y., Hirasaki, G. J., House, W.V. and Kobayashi, R. Oil and gas NMR properties: The light and heavy ends. *In SPWLA 43rd Annual Logging Symposium*. 2002. Society of Petrophysicists and Well-Log Analysts.
- [33] Gil-Vera, V.D. Viscocity Dataset, 2023, [Online]. <u>https://github.com/victorgil777/VISCOCITY/</u> <u>raw/main/DATASET.xlsx</u> (Accessed Date: September 21, 2023).
- [34] Lo, S.W., Zhang, Y., Hirasaki, G. J., House, W.V. and Kobayashi, R. Correlations of NMR relaxation time with viscosity, diffusivity, and gas/oil ratio of methane/hydrocarbon mixtures. In SPE Annual Technical Conference and Exhibition. 2000. Society of Petroleum Engineers.
- [35] Pinto, A., Ferreira, D., Neto, C., Abelha, A. and Machado, J. Data Mining to Predict Early-Stage Chronic Kidney Disease, *Procedia Comput Sci*, Vol. 177, 2020, pp. 562–567,

https://doi.org/10.1016/j.procs.2020.10.079.

- [36] Schock, C., Dumler, J. and Doepper, F. Data Acquisition, and Preparation - Enabling Data Analytics Projects within Production, in *Procedia CIRP*, Elsevier B.V., 2021, pp. 636– 640. doi: 10.1016/j.procir.2021.11.107.
- [37] Plotnikova, V., Dumas, M. and Milani, F. P. Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements, *Data Knowl Eng*, Vol. 139, pp. 102013, 2022, doi: 10.1016/j.datak.2022.102013.
- [38] Schröer, C., Kruse, F. and Gómez, J. M. A Systematic Literature Review on Applying CRISP-DM Process Model, *Procedia Comput*

Sci, Vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

- [39] Saltz, J. S. CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps, in 2021 *IEEE International Conference* on Big Data, 2021, pp. 2337–2344. doi: 10.1109/BigData52589.2021.9671634.
- [40] Solano, J. A., Lancheros Cuesta, D. J., Umaña Ibáñez, S. F. and Coronado-Hernández, J. R. Predictive models assessment based on CRISP-DM methodology for student's performance in Colombia - Saber 11 Test, *Procedia Comput Sci*, Vol. 198, pp. 512–517, 2022, doi: 10.1016/j.procs.2021.12.278.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The author developed all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The author hasn't any type of interest conflict.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0) Creative Commons Attribution License 4.0

https://creativecommons.org/licenses/by/4.0/deed.en US