IMPLEMENTATION OF A PREDICTIVE MODEL USING THE MACHINE LEARNING METHODOLOGY FOR THE LOCATION OF WATER INJECTOR WELLS IN A HETEROGENEOUS RESERVOIR

Romero Laura[1], Granobles Sebastian[2], Henao Adriana[3], Arian Sarmiento[4.]

Keywords

Machine Learning, Recover, Water injection, Heterogenous reservoir, Well placements

Abstract

The heterogeneous reservoirs are common in the world being a non-uniform and non-linear spatial distribution of rock properties such as porosity, permeability, and oil, gas, and water saturation. The petroleum industry offers limited options to predict effective locations for water injector wells in a heterogeneous reservoir. The leading country using artificial intelligence is China and it is being implemented in different sectors of daily life such as medicine and surveillance among others. The oil and gas industry use of this kind of tools is very new with a significant potential. Based upon machine learning this study predict the fluid responses approach to changes in the injector well locations, the algorithm is trained to evaluate the oil production using a series of complex patterns changing the code to make accurate decisions. The characterization of these parameters in hydrocarbon-bearing rocks is the main topic of this study, based on these properties and focusing on pairwise connectivity between water injector wells and oil producers. Using artificial intelligence, the data analysis reduces the human factor bias starting to manage a big amount of data within a shorter period of time.

Water injection is used in heterogeneous reservoirs to maintain the pressure underground and pump the fluids into the surface using the production wells, this technique can obtain around the 60% recovery of the original oil in place being an effective alternative whenever the right decisions are made. The machine learning process begins with a static and dynamic model using a specialized software to obtain the initial data from the well-known reservoir behavior of the Colombian oil basin “Caguan, Putumayo”. With this information, the algorithm XGBoost is coded in Python using some of the available libraries that allow the user a better and effective approach. The model used is called “supervised learning”; a task is given to the machine by assigning a pair of output and input for each sample given. As the name implies using training data will help the system to obtain and identify patterns that will lead to solve the mentioned task while being supervised by the user and obtaining a result that should match the data. Finally, three different scenarios this research are performed to determine the success of the process considering the production forecast.

In conclusion, the results obtained by the algorithm, compared with the running simulations, show a considerable improvement in production, showing that the predictions are useful for decision-making. In the same way, the algorithm considered some variables more relevant than others, making a comparison for each study scenario, resulting in the water saturation for scenario 1, being important when representing the amount of water that will displace the hydrocarbons within the reservoir. , the permeability for scenario 2, being vital to determine the way in which the fluids move within the pores of the rock and the location of the wells for scenario 3, being the central core of the investigation.

1. Introduction

Heterogeneous reservoirs are common in the world being a non-uniform and non-linear spatial distribution of rock properties such as porosity, permeability and saturations (oil, gas and water). The petroleum industry offers limited options to predict effective locations for water injector wells in a heterogeneous reservoir.

The Caguan-Putumayo Basin, is characterized by two different structural provinces, described as a folded zone and a geologic fault against the Colombian Andes, a relatively plane zone and stable located against the Guayana shield (Govea R & Aguilera B., 1980). The average recovery factor in Colombia is around 21% (Agencia Nacional de hidrocarburos, 2016), so it is important to improve this factor to increase the oil production in the zone, compared with different countries around the world

Secondary recovery (SR) is a strategy used in order to increase or maintain the energy of the reservoirs to improve the volume of hydrocarbons extracted when there is a decline in the natural energy of primary production. The injection of water is the most common SR in the industry due to the ease of obtaining injected fluids (sometimes the same formation water) as it physically displaces the oil. In a completely developed oil or gas field, wells may be drilled anywhere from 60 to 600m (200–2000ft) horizontally from each other, depending on the nature of the reservoir (Denis, 2019). Fluid recovery after this kind of mechanism is estimated to be between 30-50%. (Espinosa Berdugo & Torres Orellano, 2015)

It is important to assess new tools that explore different solutions regarding to conventional problems that oil and

1. laura.romero5@estudiantes.uamerica.edu.co
2. sebastian.granobles@estudiantes.uamerica.edu.co
3. henao.adriana@profesores.uamerica.edu.co
4. arian.sarmiento@profesores.uamerica.edu.co

gas industry is struggling. To reduce a large percentage of uncertainty, they need to process big data volumes that corresponds to different well and production data available, reducing the high confidence in traditional methods. Using the development of easily accessible tools, both economically and technically, obtaining satisfactory results, it is common to use specific machinery that requires special treatment to operate at full capacity. Artificial intelligence is used globally to identify patterns that could indicate possible damages before it happens giving space to predictive maintenance. In the modern world everything around us is related to a data source, machines are very effective analyzing and recognizing patterns inside it, the machine-learning model predictions make possible to obtain precise evaluation about the likely outcomes of a question, based on historical data, such as the production and injection rate. The system learns autonomously as it is fed with data, each iteration helps to reduce uncertainty and risk, using a technique called “supervised learning” that reduces human intervention to a minimum, as well as their errors. The supervised learning would automatically evaluate the optimal behavior in a particular context or environment, improving its efficiency. This type of learning is based on, a task that is given to the machine and in this research they would learn to assign the most optimal position for the injection well. As the name implies, using training data will help the system to obtain and identify patterns that will lead to solve the mentioned task while being supervised by the user and obtaining a result that should match the data (Sarker, 2021). Likewise, it would be possible to determine if some producing wells might become injectors to optimizing field production, increasing the recovery factor and the drained area.

Extreme Grandient Boosting is the most used algorithm currently of the supervised type; is a scalable, sparsity-aware and made for tree boosting, it is characterized by obtaining good prediction results more efficiently than model simulations with a more complex code. XGBoost is a useful tool to save time and resources, supports different languages and programming environments. It can also be used as a data driven tool in multiple applications, (Sampaio, Ferreira Filho, & Abelardo, 2009) applied a feed-forward neural networks as nonlinear proxies of reservoir simulation. This network can represent the multidimensional surface responses in a heterogeneous reservoir model. (Paras, Suet - Peng, Pao, & Paul, 2014) an application of Surrogate Reservoir Model (SRM) for predicting the Bottom-Hole Flowing Pressure (BHFP) at different time step for an initially under-saturated reservoir, using a SRM is based on Artificial Neural Network to regenerate the results of a numerical simulation model in considerable amount of time.

2. Methodology.

This study use data acquired in Putumayo Basin, includes general geological features of the reservoir and well models with a static and dynamic analysis. The information is analyzed and classified to use only relevant data, and dismiss repeated data with a high level of uncertainty or located outside the range of interest. Later, a part of the information is selected to be in the “learning” stage, the remainder information is left for the “implementation” stage, this part of the process is done automatically by the system without intervention of the user.

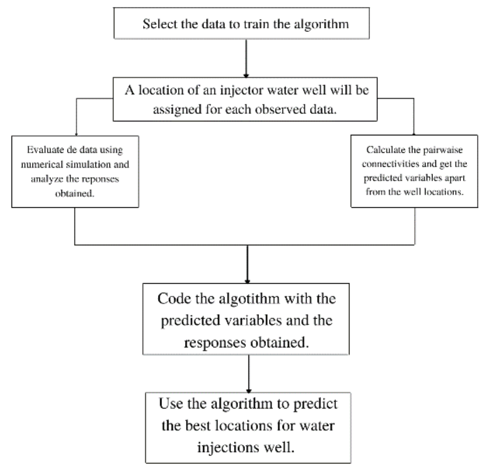

Figure 1. Workflow for the algorithm.

For injections purposes well to well, **connectivity** property is the base to determinate the behavior of hydrogeological processes in the reservoir. Generally, connectivity is not considered as a relevant factor in the reservoir, being omitted due to the importance of permeability, since this is a representation of the movement of fluids in the porous medium. Although, considering that our study focuses on the location of the wells, a variable that directly related these concepts had to be taken into account in order to obtain more precise results.

Subsequently Python software and its Visual Studio Code graphical interface must be downloaded. Specific extensions are prepared by downloading libraries using commands for their correct activation, these libraries include:

Extension	Description
XGBoost	Accurately predict a target variable
Pandas	Analyze, manipulate and organize data with rows and columns
Sklearn	Offers algorithms for classification, regression and data analysis.
Urlib	Used to work with URL’s facilating the handling
Numpy	Main library for scientific informatics
Matplotlib	Graphics generation

1. laura.romero5@estudiantes.uamerica.edu.co
2. sebastian.granobles@estudiantes.uamerica.edu.co
3. henao.adriana@profesores.uamerica.edu.co
4. arian.sarmiento@profesores.uamerica.edu.co

Table 1. Python extensions used.

After the libraries are ready, the XGBoost algorithm and the representative percentage of the data set are destined for training, the learning stage begins to identify the patterns that will lead to a prediction.

Then predictive model is implemented using the remaining percentage of the dataset to evaluate the accuracy of the learning stage and thus allow the system to work autonomously. An error analysis is carried out that allows modeling and changing some aspects that are not relevant to the project, this step is the most important, since it determines the quality of the results that will be obtained later.

Afterward the results will be obtained, analyzed and classified depending on their effectiveness to determine the success of the project, the positions found will be represented on a scale map of the reservoir that will include the proposals for the new wells established in the implementation stage.

A second simulation will be made in the input system (CMG) where the initial production forecasts will be compared against those obtained in the research and thus know if it is feasible to apply the model for decision making as a primary mechanism that turns out to be economic and convenient.

3. Calculation.

Through the research it was necessary to make some calculations, to make sure that the reservoir is heterogeneous we used the Dykstra Parsons equation, to insert the data in Python. It was necessary to make calculation of some variables such as the harmonic permeability, the fluids mobility, the operational connectivity, and the production rate.

The Dykstra Parsons equation considers an important statistical parameter to characterize heterogeneity, this coefficient varies between zero and one, with a completely uniform system having a value of zero (Arya, Hewett, Larson, & Lake, 1988).

$$V_{DP} = \frac{k_{0,5} - k_{\sigma}}{k_{0,5}} \quad (1)$$

Where V_{DP} is the heterogeneity coefficient, $K_{0,5}$ median permeability (value with 50% frequency of occurrence) and K_{σ} is t permeability at 84.1 % of the cumulative sample.

The fluids mobility is designated by the letter M and is defined as the mobility of the fluids, allowing to evaluate the effective permeability in each phase (Ferrer, 2001).

$$M = \frac{k_w/\mu_w}{k_o/\mu_o} \quad (2)$$

Where K_w is the water relative permeability, K_o is the oil relative permeability, μ_w is the water viscosity and μ_o as the oil viscosity.

To consider the permeability within the variables used for Python a calculation of an average permeability was necessary, using the values for each direction (i, j, k) a harmonic average was obtained with the following equation.

$$K_{eff} = \frac{1}{\frac{1}{K_i} + \frac{1}{K_j} + \frac{1}{K_k}} \quad (3)$$

Where K_i is the permeability in the horizontal axis, K_j represents the vertical axis and K_k is the depth axis (Bajpai, 2010).

There are four types of connectivity: geological, mechanical, operative, and dynamic. The geological connectivity is the classification of reservoirs from the structural and stratigraphic point of view using 3D seismic in the CMG software, producing layers and the presence of faults. Mechanical connectivity refers to the availability of producing layers in wells injectors and producers. For this purpose, a careful review of the history of interventions opening / isolation of each of the intervals cannonaded from each well. Although, Operational connectivity refers well connectivity observed experimentally, the field model is processed on a software, where experimental process was not made.

Using the connectivity measure presented by (Gomez-Hernandez, Carrera, & Sanchez-Vila, 2002) considering the exponent for power averaging of permeabilities.

$$K_{eff} = \left[\frac{1}{V}\int_V K(x)^c dV\right]^{\frac{1}{c}} \quad (4)$$

Where V is the volume of the cell, x is the location in space. K_{eff} is the harmonic permeability calculated before and c is the connectivity (Knudby & Carrera, 2004).

The final calculation made was for the injection rate that will affect the reservoir considering that it is directly proportional to the production rate.

$$q_{jk} = q_{jk-1} e^{-\left(\frac{\Delta t}{\tau_j}\right)} + \sum_i f_{ij} i_{ik} \left(1 - e^{-\left(\frac{\Delta t}{\tau_j}\right)}\right) \quad (5)$$

Where q_{jk} is the production rate of producer j at time k, i denotes injection rate, Δt is the time step, and τ is a time constant measure as defined in (Sayarpour, Zuluaga, Kabir, & Lake, 2007) f_{ij} is a 'gain', which is the fraction of flow from *injector i* towards *producer j* at steady state flow. This can be interpreted within this context as a measure of connectivity between injector i and producer j, because its formulation hinges solely on historical injection and production rates (Nwachukwu, Jeong, Pyrcz, & Lake, 2017)

4. Results and discussion.

The project accuracy was evaluated with three different scenarios, to determine the best locations in the

1. laura.romero5@estudiantes.uamerica.edu.co
2. sebastian.granobles@estudiantes.uamerica.edu.co
3. henao.adriana@profesores.uamerica.edu.co
4. arian.sarmiento@profesores.uamerica.edu.co

reservoir. These scenarios were considered with the number of available wells in a determined moment of time and as a result, the number of wells (input data) were, 9, 12 and 15. Each case was evaluated with values of time, feature score, and the error.

Training data	R^2 (Training data)	R^2 (Testing data)
9 wells	0.999	-2.049
12 wells	0.999	0.964
15 wells	0.999	0.938

Table 2. Scenario performance

The accuracy is determined depending on the difference between each value of R^2, the higher this difference is, higher will be the uncertainty for each case.

The model used in Python will consider a different set of variables that are relevant, these features will be determined automatically by the system and change for each case.

Scenario number 1: In this scenario the training data were 9 wells, and the most important feature is the initial water saturation (Figure 2), considering this variable is the fraction of the pore volume of water compared with the available space within the rock total pore volume. Assuming this volume is either filled with water, oil or gas, water saturation must be measured or estimated for the reservoir characterization and it is the most challenging petrophysical calculation for estimating hydrocarbon-in-place (Cui, 2015).

To estimate the optimal locations the algorithm took 2,59 seconds,as a result, we obtained the coordinate of a single injector well represented in figure 3 using the CMG simulator, the injection rate was obtained as well, in this case 238 BBL/day.

Figure 2. Feature importance score, Scenario 1

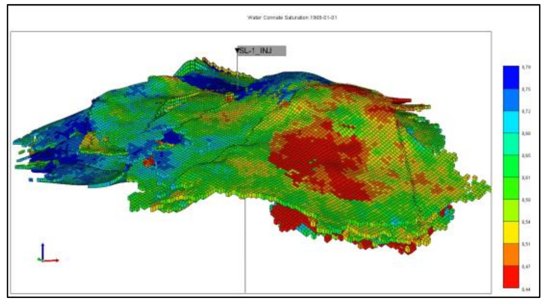

Figure 3. Well location Scenario 1

Analyzing the data results it is possible to see that the testing data tendency (R^2) compared with the training data tendency (R^2) have the biggest difference between each other, within all the scenarios. Comparing the initial simulation with the model, where the respective drilled well is already located in figure 3, it is not possible to see a representative difference in the prediction of production from the reservoir as it is shown in the figure 4, the blue line represents the original production without the new well, by the other hand, the red line is the new production with the injection well.

The prediction for 9 wells is lower than the original file as there is less wells to consider, this was made taking in consideration that some wells were too old to be considered in the new prediction.

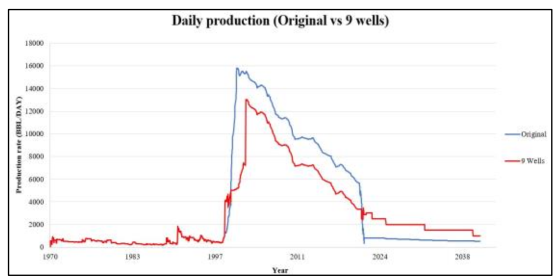

Figure 4. Oil production (9 wells Vs Original model)

Scenario number 2: In this scenario the training data were 12 wells, and the most important feature is the harmonic permeability (Figure 5), this value is important in the injection models as it can be used to predict the well productivity index (Law, 1943), this represents the capacity of a rock to transmit the fluids (Berg, 1970).

To estimate the optimal locations the algorithm took 2,70 seconds being the scenario that took the longest time to be completed, as a result we obtain the location of two injector wells represented in figure 6 using the CMG simulator, the injection rate was obtained as well, in this case 238 and 317 BBL/day.

1. laura.romero5@estudiantes.uamerica.edu.co
2. sebastian.granobles@estudiantes.uamerica.edu.co
3. henao.adriana@profesores.uamerica.edu.co
4. arian.sarmiento@profesores.uamerica.edu.co

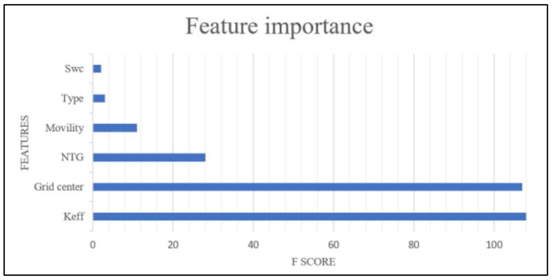

Figure 5. Feature importance score, Scenario 2

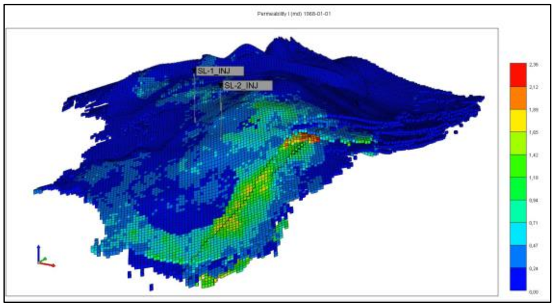

Figure 6. Well locations Scenario 2

Analyzing the data results it is possible to see that the testing data tendency (R^2) compared with the training data tendency (R^2) have the smallest difference among all the scenarios. Comparing the initial simulation with the model where the respective drilled wells are already located in figure 6, it is possible to see the best kind of prediction of production from the reservoir as it is shown in the figure 7, the blue line represents the original production without the new wells and, the red line is the new production with the injection well.

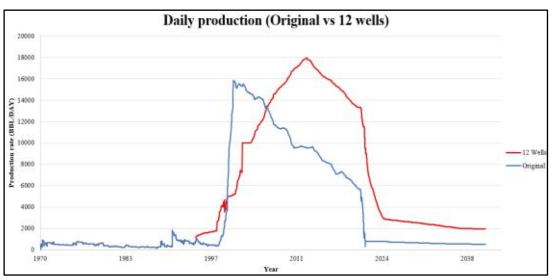

Figure 7. Oil production (12 wells Vs Original model)

Scenario number 3: In this scenario the training data were the 15 original wells, and the most important feature is the grid center which represents the location in space for each well, this value is vital in the injection model as it represents the place where the next wells will be drilled.

To estimate the optimal locations the algorithm took 2,53 seconds as a result we obtain the location of three injector wells represented in figure 9 using the CMG simulator, the injection rates were obtained, in this case 950, 634 and 238 BBL/day.

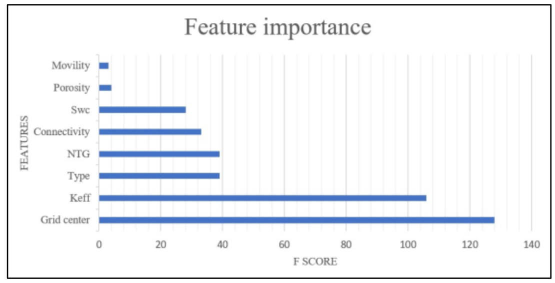

Figure 8. Feature importance score, Scenario 3

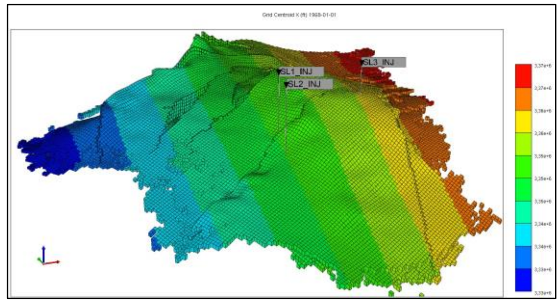

Figure 9. Well locations Scenario 3

Analyzing the data results it is possible to see that the testing data tendency (R^2) compared with the training data tendency (R^2) have an average difference among all the scenarios. Comparing the initial simulation with the model where the respective drilled wells are already located in figure 9. Although, the prediction of this scenario is not the most optimal, we can see that in the oil production graph has the highest production rates compared to the other models evaluated in the software. Nevertheless, we see that it declines very quickly, decreasing available production for future years, as it is shown in the figure 10, the blue line represents the original production without the new wells andthe red line is the new production with the injection well.

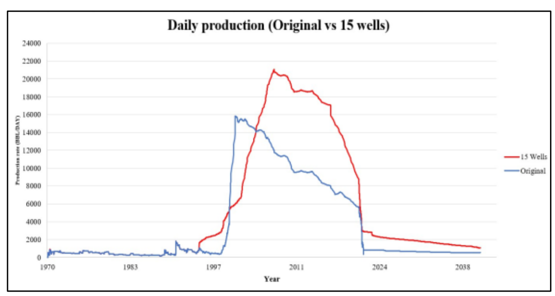

Figure 10. Oil production (15 wells Vs. Oringinal model)

5. Conclusions

This research considered the most relevant petrophysical variables of a reservoir to obtain accurate results when using proxy models, these variables are key to build an algorithm that analyzes the response of reservoirs to water injection. The mobility equation studies the behavior of fluid within the rock pores, in this case water and hydrocarbons, the connectivity values

1. laura.romero5@estudiantes.uamerica.edu.co
2. sebastian.granobles@estudiantes.uamerica.edu.co
3. henao.adriana@profesores.uamerica.edu.co
4. arian.sarmiento@profesores.uamerica.edu.co

that represent how easy it is for the fluids to move inside the reservoir from one well to another (well to well connectivity). In the same way, the heterogeneity of the reservoir was confirmed to be certain that the study area is adequate, using the Dysktra parsons' method, obtaining a factor of 0,67.

Despite considering variables as important as porosity and connectivity, for the algorithm it was not possible to recognize patterns related to production as it had such small variances between its data, obtaining a feature importance score given by the program of 4 and 32 respectively only for Scenario 3.

An alternative to conventionally methodology is proposed for the prediction of injection well locations in heterogeneous reservoirs. This method demonstrated that it is possible to save time and money in decision-making procedures that involves a high uncertainty. Using the Extreme Gradient Boosting algorithm, the best locations for injection wells in a heterogeneous reservoir were predicted, using known values that influence hydrocarbon production. During the training process, the algorithm calculates the deviation percentages and returns an approximate error result, allowing to increase the reliability of the model.

It is concluded that the most optimal scenario is number 2, considering that the average production of 7,900 BBL/day is relatively constant over time. The location of the wells in scenario 2 turns out to be optimal as they are close to the fault where this natural channel is used for fluid migration. By having a single location of an injector well in scenario 1, the changes in the average daily production of 4900 BBL / day are not significant compared to the other scenarios and even with the original model, showing that a single injector well is not in the ability to push the fluids in the area. For scenario 3 despite having the maximum production in a day of 21000 BBL/day, it's average production is low compared to scenario 2, showing a quick decline.

For scenario 1, the most important variable was the initial water saturation with an F score of 50 because this defines the initial amount of fluid that will be pushed. For scenario 2, the most important variable was harmonic permeability F score of 108, considering that this represents the ability of fluids to move within the pores of the rock. For scenario 3, the most important variable was the location of the wells F score of 128, demonstrating the importance of their distribution in the reservoir to ensure a good recovery factor.

6. References

Agencia Nacional de hidrocarburos. (2016). CONVOCATORIA PARA LA EJECUCIÓN DE PROYECTOS DE I+D+i EN RECOBRO MEJORADO DE HIDROCARBUROS (EOR). DEPARTAMENTO ADMINISTRATIVO DE CIENCIA, TECNOLOGIA E INNOVACIÓN - COLCIENCIAS -.

APD, R. (4 de marzo de 2019). ¿Qué es Machine Kearning y cómo funciona? Obtenido de APD España: https://www.apd.es/que-es-machine-learning/

Arya, A., Hewett, T. A., Larson, R. G., & Lake, L. W. (1988). Dispersion and Reservoir Heterogeneity. SPE Reservoir Engineering.

Bajpai, N. (2010). Business Statistics. Delhi: dorling kindersley.

Berg, R. R. (1970). Method for Determining Permeability from Reservoir Rock Properties. Gulf Coast Association of Geological Societies Transactions, 303-317.

Cui, E. x. (2015). Initial Water Saturation Calculation through Integrated Methods for 3D Unconventional Reservoir Modeling. SPE Kuwait Oils & Gas Show and conference, 2.

Data Robot. (s.f. de s.f. de s.f.). Data Robot. Obtenido de Data Robot, Inc.: https://www.apd.es/que-es-machine-learning/

Denis, P. N. (2019). Handbook of Fire and Explosion Protection Engineering Principles for Oil, Gas, Chemical, and Related Facilities. Milwaukie, OR: 32-50.

Espinosa Berdugo, C. A., & Torres Orellano, K. J. (2015). Técnicas de recobro y recobro mejorado en yacimientos con crudos livianos, pesados y extrapesados. Ingenierías USB Cartagena, 5-6.

Feldstein, S. (2019). The Global Expansion of AI Surveillance. Washington, DC: Carnegie Endowment for International Peace.

Ferrer, M. P. (2001). Inyección de agua y gas en yacimientos petrolíferos. Maracaibo: Ediciones Astro Data S.A.

Gomez-Hernandez, J., Carrera, J., & Sanchez-Vila, X. (2002). Evaluation of Different Measures of Flow and Transport Connectivity of Geologic Media. Barcelona: KLUWER.

Govea R, C., & Aguilera B., H. (1980). Geología de la cuenca del Putumayo. Bucaramanga, CO.

Knudby, C., & Carrera, J. (2004). Evaluation of Different Measures of Flow and Transport Connectivity of Geologic Media. geoENV IV — Geostatistics for Environmental Applications pp 235-246, 235–246.

Law, J. (1943). A Statistical Approach to the Interstitial Heterogeneity of Sand Reservoirs. Society of Petroleum Engineering.

Nguyen, A. P., Lasdon, L. S., Lake, L. W., & Edgar, T. F. (2011). Capacitance Resistive Model Application

1. laura.romero5@estudiantes.uamerica.edu.co
2. sebastian.granobles@estudiantes.uamerica.edu.co
3. henao.adriana@profesores.uamerica.edu.co
4. arian.sarmiento@profesores.uamerica.edu.co

to Optimize Waterflood in a West Texas Field. SPE Annual Technical Conference and Exhibition.

Nwachukwu, A., Jeong, H., Pyrcz, M., & Lake, L. W. (2017). Fast evaluation of well placements in heterogeneous reservoir models using. Journal of Petroleum Science and Engineering.

Paras, M. Q., Suet - Peng, Y., Pao, W., & Paul, S. J. (2014). Surrogate reservoir modeling-prediction of Bottom-Hole Flowing Pressure using Radial Basis Neural Network. IEEE Xplore.

Sampaio, T. P., Ferreira Filho, V. J., & Abelardo, D. (2009). An Application of Feed Forward Neural Network as Nonlinear Proxies for Use During the History Matching Phase. Latin American and Caribbean Petroleum Engineering Conference.

Sarker, I. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. Springler Link.

Sayarpour, M., Zuluaga, E., Kabir, S., & Lake, W. L. (2007). The Use of Capacitance-Resistive Models for Rapid Estimation of Waterflood Performance. Annual Technical Conference and Exhibition.

1. laura.romero5@estudiantes.uamerica.edu.co
2. sebastian.granobles@estudiantes.uamerica.edu.co
3. henao.adriana@profesores.uamerica.edu.co
4. arian.sarmiento@profesores.uamerica.edu.co