Petroleum and Coal

Article **Open Access**

Machine Learning Approach to Rock Properties Determination using NMR Log Data

Waleed M. Osman[1,2], Abdelfattah A. Rashid[1], Adel M. Salem[2]

[1] Faculty of Energy & Environmental Engineering, The British University in Egypt (BUE), Egypt
[2] Faculty of Petroleum & Mining Engineering, Suez University, Suez, Egypt

Received April 26, 2021; Accepted October 8, 2021

Abstract

Recently machine learning applications have been widely used in oil and gas industry. Artificial intelligence applications in petrophysics are used for predicting some petrophysical parameters such as porosity and permeability. In this paper we are calculating petrophysical parameters in some field in the Middle East Gulf area and use these calculated parameters applying machine learning KNearest Neighbor (KNN) regression method to predict these parameters in some other portions of the field. The data used in this study were Nuclear Magnetic Resonance (NMR) logs. The constructed prediction models estimate porosity and permeability logs using a mathematical model of machine learning (KNN) method followed by preprocessing data analysis as importing libraries, datasets, processing dummy variables and feature scaling. NMR porosity and permeability logs were taken as dependent variables (Y) depending on independent input variables (X) such as T2 distribution and T2LM that resulted in essential prediction of porosity and permeability at a certain depth interval of interest. The used KNN regression models showed a valid accuracy of the numerically predicted logs.

Keywords: *Machine Learning (ML) CMR; NMR; TCMR; KSDR; KNearest Neighbor (KNN).*

1. Introduction

During 2010-2020, the world witnessed unprecedented real-world applications of big data analytics, machine learning (ML), and artificial intelligence (AI) in many fields including oil industry and autonomous vehicles. A breakthrough in big data and AI application for human life development, can operate as a good asset among applications sets [1-3]. Not surprisingly, digital change, big data, and ML/AI have become highlighted words in the oil industry with many field applications reported from both engineering and geosciences majors. Since 2016, almost all conferences have dedicated plenary or technical sessions related to such themes. A considerable number of articles based on the technical progresses of big data analytics in the upstream oil industry can be demonstrated in the industrial journals such as Journal of Petroleum Technology and the Leading Edge. Also, journals of scholastics have performed special paperwork to those edge fields that were applied in petroleum engineering or geosciences [4-6].

By investigating the upstream applications of ML/AI in the literature, petroleum engineers and geoscientists have been using machines and/or computers to use their databases for period of time now to formulate complex tasks. (Figure 1) shows the trend for published papers on ML/AI fields in digital libraries. The AI applications peak can be dated 30 years back when old-fashioned generation predicting codes such as linear regression, genetic algorithms, and neural networks started to vector their momentum towards oil industry. One may notice that the number of ML researches surpassed the number of AI researches near the year of 2015, which may assure that ML has become a concentrated area of interest than AI. Deep learning criteria revitalized the ML/AI usages in many other industries and started another wave of AI applications in the petroleum industry in the past years [7-8].

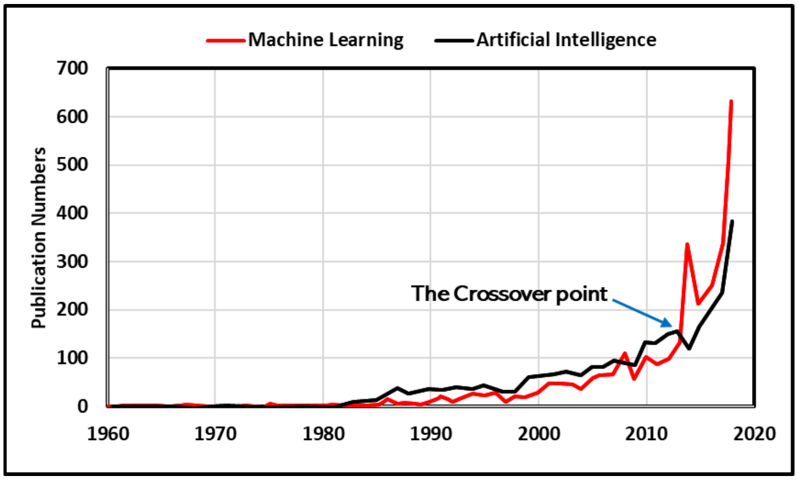

Figure 1. Illustrations of number papers on ML/AI in the OnePetro digital library since 1980 [9]

Important investments from both reserves owners and service companies are in place to mount this digital transformation revolution. It is hopefully known that the technical outbreak in data science analytics and AI fields will start more impacts in the next years. Many operators have established specialized groups or teams working on ML/AI Data Analytics. Also, lots of high-tech new investments focusing on ML/AI Data Analytics solution products have started around the world. ML, as a critical component has huge value in mining the business value grasped from the streaming petrophysics big data. Knowing of this value and a high-level overview of data types, codes, and the field of applications existence in different technical fields will surely benefit the workers of such industry in knowing machine learning codes and algorithms to their own database to solve serious field challenges.[9-10]

Also many slides of logging topic is discussed which are in general: the lithology condition and the way that the nature use to form the rock containing the petroleum matter and the rock used to seal or prevent this matter from escape. Also the logs used to identify many concepts in order to illustrate ways to estimate the hydrocarbon reserve deep down the surface and its pay zone and those logs are considered to be resistivity logs (and its measuring principles), gamma ray logs, spectral gamma ray logs, porosity logs, sonic porosity logs, density logs and neutron tools. Then the tools which are being used to determine such important data which is CMR which is the main output used in this demonstration of estimating porosity and permeability using machine learning [9-10].

2. Methodology

2.1. ML Algorithms Overview

ML is extracting patterns and formation structures from historical offset data (training data for supervised or unsupervised learning) by interfering with a ML workflow (reinforcement learning) to fully characterize, identify, or even predict the behavior of the system. Also, systems could be massively including thousands of attributes. ML coded-algorithms codes use the data to learn the behavior of the process system without prior acquisitioned knowledge of the nature of environments between data measured points. ML is also well suited to appoint those challenges where theoretical physics-based acknowledgement is still not complete so that we do not have a reasonable number of measurements, estimations and observations, variable ML techniques are broadly classified into supervised learning and unsupervised learning [11].

Two commonly used supervised learning tasks are linear regression and K-means classification. Linear regression describes the linear relation between two variables on a Cartesian graph to be used then for numerical predictions, K Nearest Neighbor (KNN) is the same as linear regression method but not only for regression but also for classification processing, also. These methods are means to develop a prediction mean of X and Y variables to build estimation numerical models, identifying the variability of data distribution via cartesian graphs through a preprocessing data analysis of training this new numerical model using a fraction of this dataset, as an adjustment the data for a graphical relation between the X and Y variables to estimate brand new values of unused data in the future using the same trained numerical model. As an example, regression methods generate continuous outputs/responses (e.g., phi, Sw, K), on the other hand classification methods generate discrete output/classes/labels (e.g., rock typing, facies) [11].

Unsupervised learning processes attributes to identify commonalities to learn the relationships and data patterns in the unsupervised data. As unsupervised models are considered to be numerical models that needed no prior training in the data preprocessing segment of the numerical model building. As an example, clustering, anomaly detection, dimensionality reduction, feature ranking, and data visualization are few duties based totally on unsupervised learning. In supervised and unsupervised learning, a comprehensive 'training dataset' is constructed which covers as much of the dataset system parameters missing data as possible. For supervised learning, an arbitrary set of the data is put along with a completely independent test [11].

2.1.1. Supervised Learning VS Unsupervised Learning

In supervised learning, data-driven model building is done by processing a labeled database that involves required attributes and responses. Supervised learning fitches the processing model that generates the process outputs based on the chosen inputs. A physics-driven modeling is a theory of mathematical contouring that relates chosen inputs with desired output.

Supervised learning acknowledges patterns in the available data, learns from data observatories, and decides the necessary estimations based on statistical contouring of chosen inputs and desired outputs. As the process is going on building the supervised learning simulator model, the estimations are compared to the desired output and the model is developed based on a function loss. This process goes forwards till the data-driven simulator model demonstrates a high level of precision and performance so the loss function can be lowered.

Unsupervised learning processes databases to identify estimated patterns, useful relations, and commonalities without using offset data, labels, and human restricted instructions. This ML methodology appoints the dataset towards a certain way that illustrates the code system structure, variance, density, and distribution. The ML dataset might mean a group the data values clustered or arrange it into a way that is more organized or easy visualized.[11]

2.1.2 Regression VS classification

Supervised learning analytics training dataset and products function as it can contour new data examples. Based on the type of the desired output, supervised coded algorithms are categorized into regression ML and classification ML.

Regression is generating continuous desired outputs (e.g., porosity, saturation, permeability). It has the rule of determining the contribution and correlation of attributes that produce a certain output. Classification ML methodology, on the other hand, generates discrete output/classes/labels (e.g., lithology, facies, and rock types). Unlike regression, classifications rule is done when a certain output is allocated into categories and the classifier reports a label/class based on certain attributes contains a list of some commonly used ML coded algorithms in the petrophysics methodology and this is a complete list and coded algorithms can be adapted to fit for different purposes [11].

2.2. CMR Porosity and permeability measurement

Porosity is a normal triple combo log. [12-13], affected by rock grains distribution and consolidation material [14].. Starting with porosity measurement, TCMR is considered to be the total porosity value calculated from CMR (Combinable Magnetic Resonance) T2 Distribution application that resulted from CPMG echo-train, the echo train signal is a multi-exponent decay function that draws a curve which is mathematically can be stated in multiple decay time constants, also these decay time constants provide a T2 relaxation distributed curve as shown (Figure 2) [15].

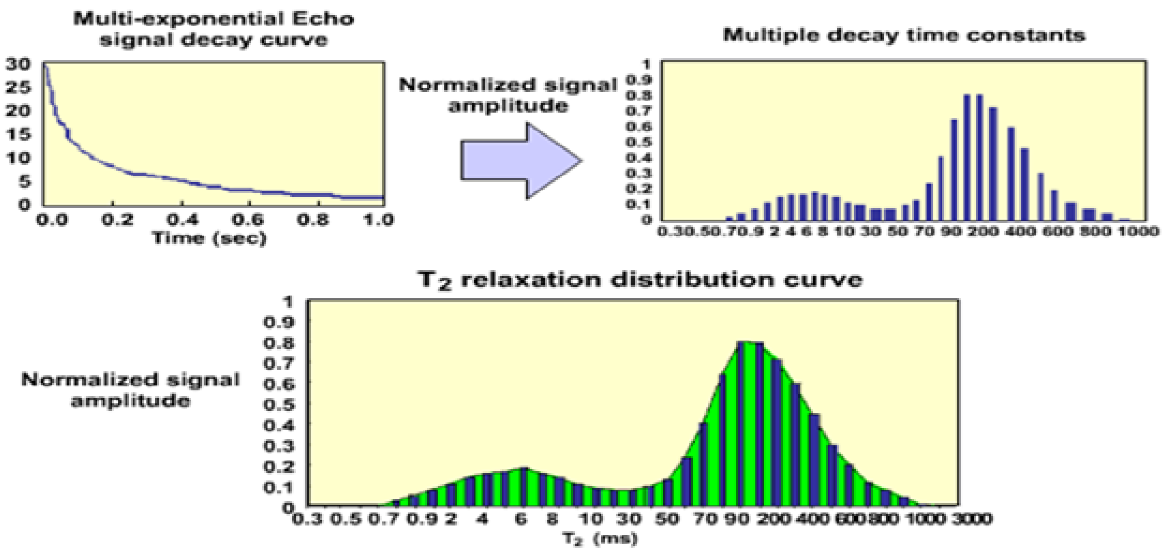

Figure 2. Echo-train and T2 distribution cuvres [15]

By integrating the T2 = distribution curve numerically, TCMR is now obtained, with the T2 log mean (T2LM), CMR porosity (TCMR), Schlumberger Doll Research Permeability log (KSDR) can be petrophysically calculated easily from (Equation 1). [16-17]

$$KSDR = A^{*}\,(TCMR)^{B}{*}(T2LM)^{C} \qquad \text{[Eq 1]}$$

where A, B and C are constants equal to 2, 2 and 4 respectively.

2.3. TCMR and KSDR numerical prediction using KNN

When a certain dataset that may encounter non-linear potential, k-Nearest Neighbors (k-NN) is an algorithm that is used for making predictions/predictions, k-NN examines the dataset variables' values of the points around it (neighbors) to determine the value of the point of interest. It is considered to be an averaging method that majorly depends on the Cartesian distance range between the data points as it relates the numerical value of each point with its coordinates, by measuring the difference between the Cartesian distances and numerical values, precise values can be estimated in between. A simple formula is basically a formula where a trend line or a slope of a line on x and y axes. By going through this procedures, variables and coefficients as Y the dependent variable (Y) which is depending on independent variable (X) can be estimated, for instance, KNN machine learning method can be used for such a petrophysical study applied numerically by (Equation 2).

KNN regressor is a ML method to predict numerical variables by assigning them to the most similar regressor algorithm labeled examples. The essential concept that actually KNN algorithm is all about is the distance between two data points on a Cartesian graph, such measurement is employed inside the KNN algorithm as Euclidean distance which can be mathematically calculated by [Equation 2] [18].

$$D(p,q)=\sqrt{(p_1-q_1)^2+(p_2-q_2)^2+\cdots+(p_n-q_n)^2}$$

where p and q are subjects (coordinates) to be put into a comparison with another data points to evaluate the variability of the numbers regarding the distance between them.

3. Results

3.1. NMR/CMR Petrophyscial Case Study

(Figure 3) shows an actual case study from the Middle East, visualizing a triple combo logs alongside with wet lithology column and CMR T2 distribution petrophysically calculated logs. Within the depth interval shown in (Figure 3), a pay zone was evaluated petrophysically, using the triple combo data, a virgin depth of investigation wet lithology column was computed showing an economic amount of hydrocarbon, also NMR T2 distribution log in the sixth log track shows a potential of a free fluid accumulation approximately within the whole depth interval of the productive zone. TCMR and KSDR were computed petrophysically using the T2 distribution log as shown in the seventh and eighth tracks respectively in (Figures 3).

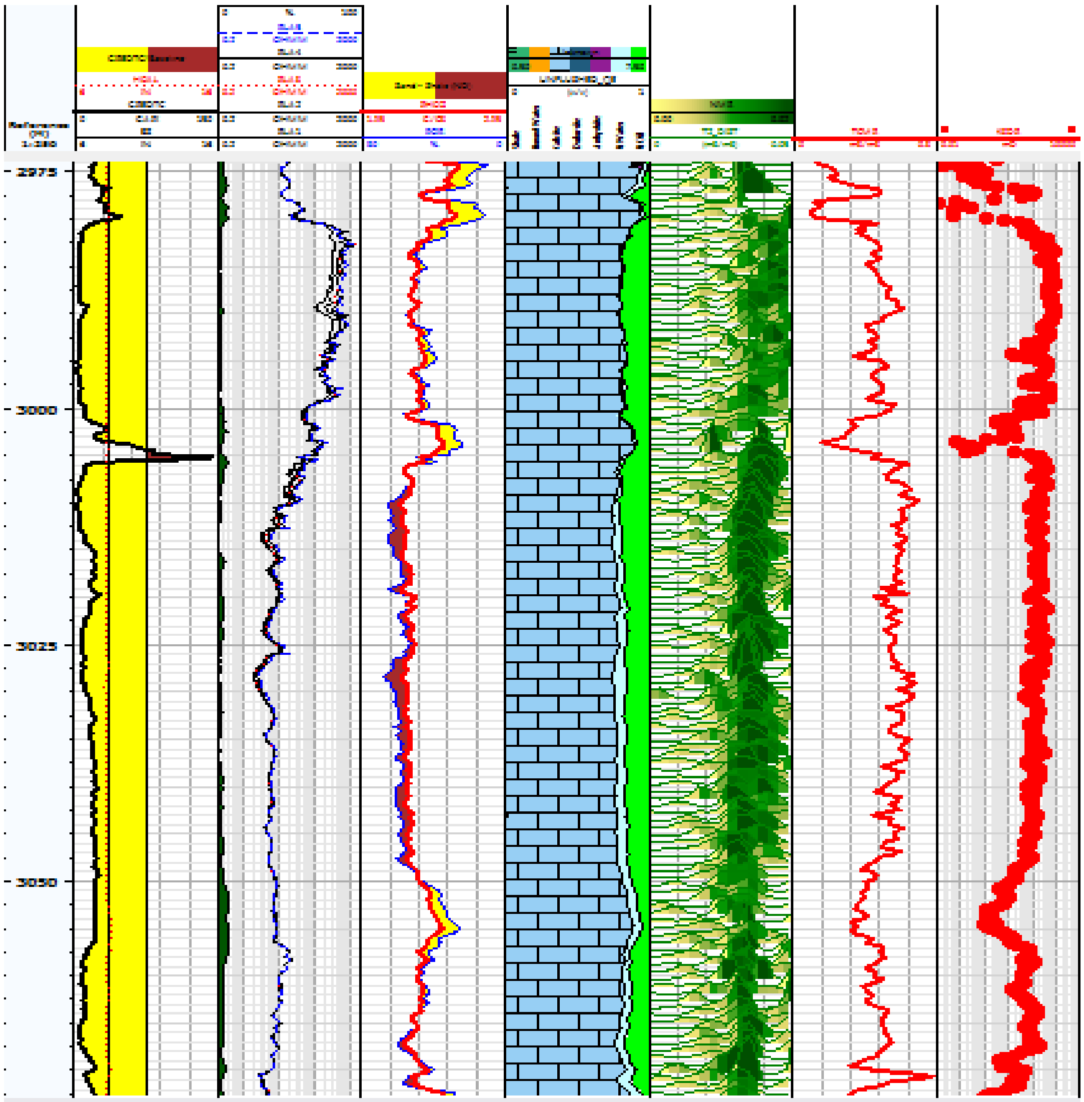

Figure 3 Triple combo, wet lithology column and T2 distribution curves

3.2. TCMR and KSDR numerical prediction using KNN

In this case study, KNN algorithm is applied upon petrophysical NMR/CMR T2 distribution and T2LM as X variables to predict TCMR and KSDR which are Y variables. Characteristics of the algorithm is collected from training dataset to setup a regressor model for each of porosity (TCMR) and permeability (KSDR) variables outputs. KNN is considered to be a distance dependent ML algorithm, a dataset file is built with the X variables and Y variable first to train the KNN model. As the code learns further, more accuracy is obtained from the surrounding wells having CMR wireline jobs were run into it. 6000 points dataset file was used to train the model before it was used to predict TCMR and KSDR in the well shown in (Figure3). As a considerable score was achieved from that training process, now the model can be used to predict the dependent variables (TCMR and KSDR) inside the targeted well as shown in (Figures 4, 5, and 6).

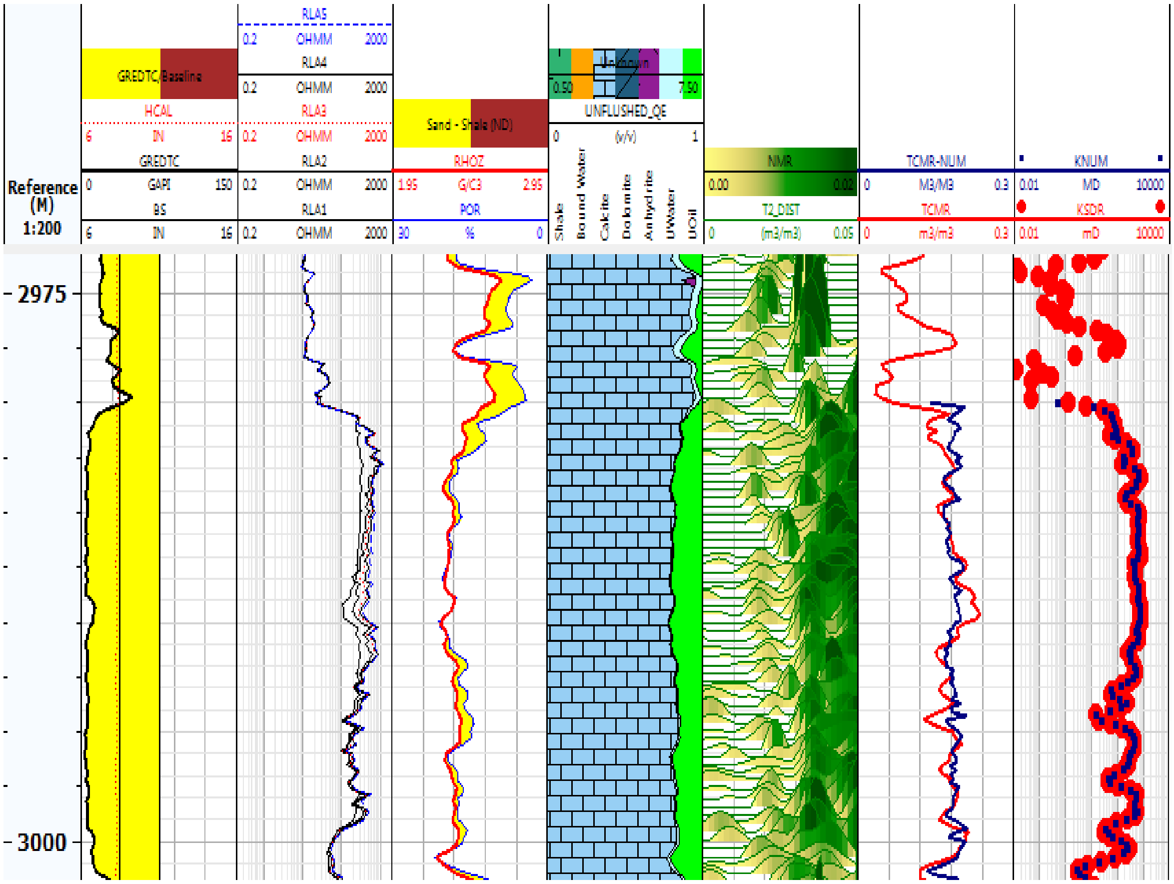

Figure 4. Zone 1 Triple combo, wet lithology column, T2 Distribution cuvres and petrophyscial / numerical porosity and permeability calculations

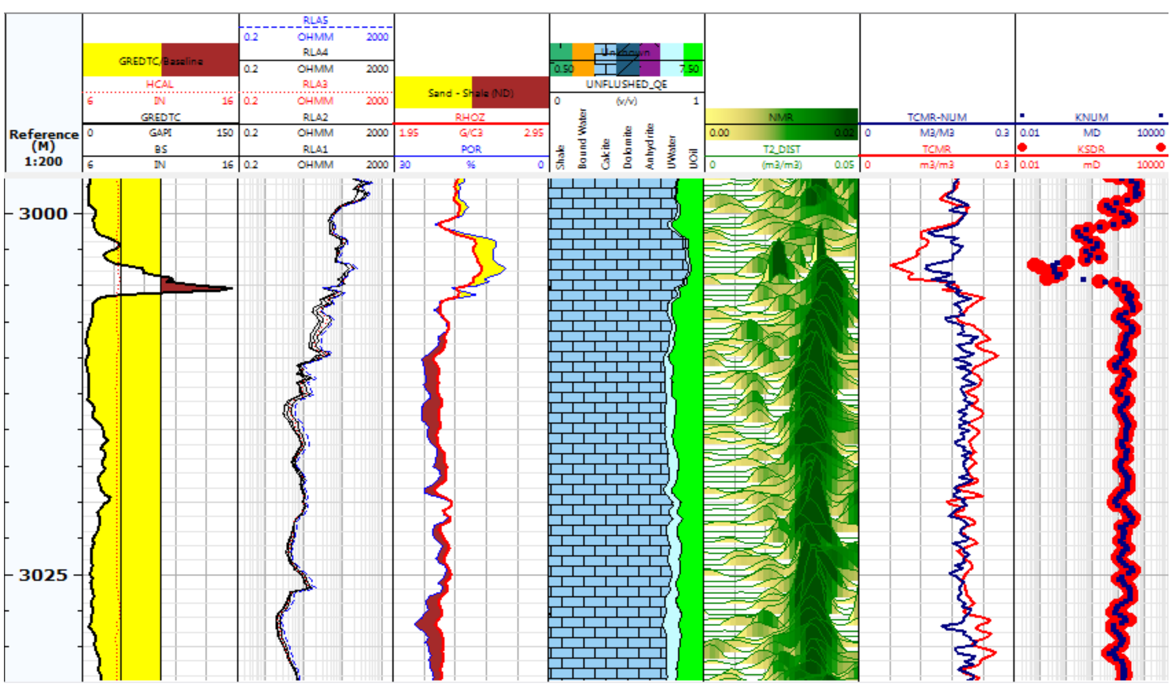

Figure 5. Zone 2 Triple combo, wet lithology column, T2 Distribution cuvres and petrophyscial / numerical porosity and permeability calculations

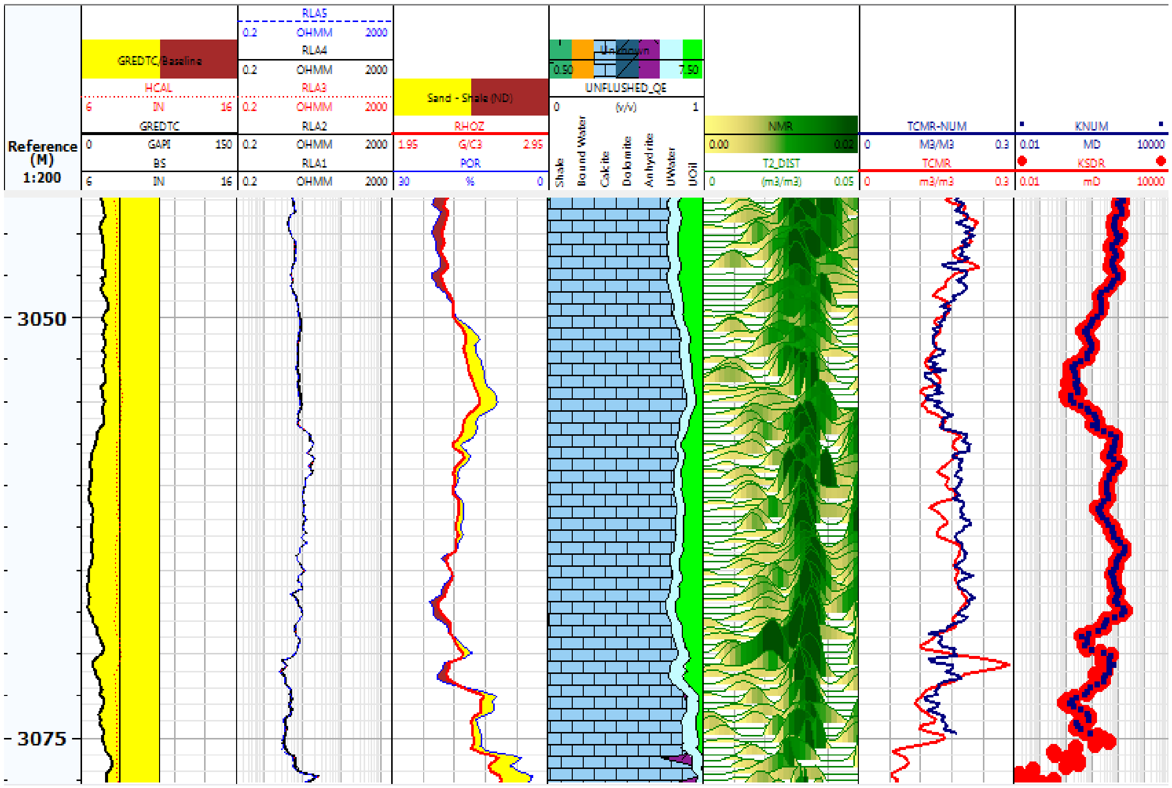

Figure 6. Zone 3 Triple combo, wet lithology column, T2 Distribution cuvres and petrophyscial/numerical porosity and permeability calculations

As shown in (Figures. 4, 5, and 6), the accuracy of the of the numerically predicted logs (blue curves in log track 7 & 8) porosity (TCMR-NUM) and permeability (KNUM) are considered to be precise if compared to the actual CMR petrophysical porosity and permeability logs. T2-distribution and T2LM were used to be the X variables for the training and validation for TCMR-NUM (Y-variable) prediction. TCMR and T2LM were also used to be the X- variables for training and validation processes for KNUM (Y variable). However the python TCMR-NUM and KNUM models showed very poor scores so feature scaling was applied on both models which resulted in approximate scores allocated to be 89% for KNUM model and 85% for TCMR-NUM in average, thus certain deviations are located at some depths due to the mathematical error that occurs in the ML numerical prediction studies.

4. Conclusion

This paper illustrated a numerical study of porosity and permeability using the well logging datasets of CMR/NMR applications and data analytics synchronization of it using machine learning KNearest Neighbor (KNN) method reviewing the literature of machine learning and Petrophysical Data Driven Analytics (PDDA), big data definition. Running the wireline unconventional CMR/NMR logs, a numerical prediction based upon petrophysical data of 2 wells were obtained to be used as training data for a third well porosity and permeability prediction. An estimation of porosity and permeability were predicted using mathematical model of machine learning (ML) method which was KNN for regression preceded by the preprocessing data analytics represented by model X variables feature scaling process and model training. Finally, an acquisition of numerical TCMR and KSDR predictions were obtained proving a promising precision when compared to the petrophysical calculated CMR porosity (TCMR) and permeability (KSDR) logs.

References

[1] Baruch J. Steer driverless cars towards full automation. Nature, 2016; 536: 127.
[2] Berger C. A competition for self-driving miniature cars to a standardized experimental platform: concept, models, architecture, and evaluation. Journal of Software Engineering for Robotics, 2014; 5(1): 63-79.
[3] Magdy KM, Gawish AA, Salem AM. Novel Applications of Artificial Intelligence Neural Network in Hydraulic Fracturing. International Journal Of Scientific & Technology Research, 2020; 9(12): 47-53.
[4] Mehta A. Tapping the value from big data analytics. Journal of Petroleum Technology, 2016; 68(12): 40-41.
[5] Hall B. Facies classification using machine learning. The Leading Edge, 2016; 35(10): 906–909.
[6] Carpenter C. Geology-driven estimated-ultimate-recovery prediction with deep learning. Society of Petroleum Engineers, 2016; 68(05): 74-75.
[7] Jacobs T. Pioneer's analytics project reveals the good and bad of machine learning. Journal of Petroleum Technology, 2018; 70(9).
[8] Ma SM. Technology Focus: Formation Evaluation. Society of Petroleum Engineers. (2018). J Pet Technol., 70(08): 50.
[9] Xu C, Siddharth M, Srinivasan P, and Ma S. When Petrophysics Meets Big Data: What can Machine Do?. SPE Middle East Oil and Gas Show and Conference. Manama, Bahrain. 2019. https://doi.org/10.2118/195068-MS
[10] Saputelli L. Technology focus: Petroleum data analytics. Journal of Petroleum Technology, 2018; 70(10).
[11] Xu C, and Misra S.: Petrophysics Data-Driven Analytics: Theories and Applications. SPWLA 2018 STC Panel Discussions Summary. April 16-17, Houston TX. SPWLA Newsletter September 2018.
[12] Osman W, Kassab M, El Gibaly A, and Samir H. Petrophysical evaluation of sandstone gas reservoir using integrated well logs and core data for the Lower Cretaceous Kharita formation, Western Desert, Egypt. Journal of petroleum Exploration and Production Technology. 2021; 11(10): 3723-3746.
[13] El sayed AG, Kassab M, and Osman W. Evaluation of Petrophysical and Hydrocarbon Potentiality for the Nuba A, Ras Budran oil field, Gulf of Suez, Egypt .2021. Egyptian Journal of Chemistry, 2021; 64(7): 3387-3404.
[14] Al rahmani BA, Al shibli KA, Attia AM. Influence of grain size and consolidation pressure on porosity of rocks. Geotechnical Special Publication, 2005; (130-142): 2445-2457.
[15] Shell/Schlumberger. Petrophysics Distance Learning:. Hague, Netherlands: (1999).
[16] Coates GR, Gardner JS, and Miller DL. Applying pulse echo NMR to shaly sand formation evaluation. Transactions of the SPWLA 35th Annual Logging Symposium, Tulsa, Oklahoma, USA (1994), paper B.
[17] Kenyon W, Day P, Straley C, Willemsen J. A Three-Part Study of NMR Longitudinal Relaxation Properties of Water-Saturated Sandstones. SPE Formation Evaluation, 1988; 3(03):622–636.
[18] Lantz B. Machine Learning with R: Expert techniques for predictive modeling to solve all your data analysis problems. 2nd ed. Birmingham: Packt Publishing; 2015: 426 p.

To whom correspondence should be addressed: Dr. Adel M. Salem , Petroleum Engineering Department, Faculty of Petroleum & Mining Engineering, Suez University, Suez, Egypt, E-mail: adelmsalem@yahoo.com