Saving Grace of Unsupervised K-Means Clustering of Facies Classification Using Well Log Measurements

Elnara Mammadova
School of Mathematics, Computer Science and Engineering
City, University of London

***Abstract*—In this study, the focus is understanding the subsurface geology present at three well locations in block 15/9 in Volve Field, located in the Norwegian North Sea, by use of a variety of electrical measurements generated from well logging technologies. Petrophysical data analysis were applied in an integrated approach, coupled with thorough data cleaning and outlier removal, to study the log (curve) responses of the wells and asses the feasibility of distinguishing different lithological trends. Specifically, unsupervised cluster analysis using K-Means Clustering algorithm were used to group well log measurements into distinct lithological groupings, known as facies.**

Keywords—unsupervised, clustering, K-Means, facies, lithology

I. INTRODUCTION

Since the global crude oil price slump began in June 2014, companies in the oil and gas industry had to reduce their well delivery costs in a bid to increase their profit margin. These cost reduction initiatives in a low oil price regime meant less data acquisition, which subsequently meant higher uncertainty in hydrocarbon predictions. Consequently, the industry has slowly shifted its focus from conventional geoscientific methods such as manually assigning lithofacies by human interpreters which is a very laborious and time-consuming process, often resulting in noisy data. As an alternative, a more data science driven approaches are being employed to reduce uncertainties and predict more accurate lithology trends without the associated costs [1].

The ideal sources for lithofacies classification are core samples of rocks extracted from wells. Nevertheless, amassing actual lithological samples is economically unfavourable and time-consuming. Therefore, a method for classifying facies from indirect electrical measurements from well logs is essential.

The current study demonstrates the application of unsupervised cluster analysis on petrophysical properties calculated from wireline logging measurements from three wells in Equinor's Volve field located at the Norwegian Continental Shelf (NCS). The results were later compared with the conventional cuttings' descriptions acquired during mudlogging for quality assessment.

II. ANALYTICAL QUESTIONS

This study aims to address frequent questions faced by the oil and gas industry, where working with noisy and missing data has become the norm. There are two aspects to this study. The first aspect includes the use of data science methodologies in removing noise and outliers from available dataset and prediction of missing log measurements by use of supervised learning. Later an unsupervised clustering method is employed to predict facies classification based on the derived porosities, permeabilities and shale volumes. Below are the three main questions we aim to answer by the end of our research.

1. How effective can we be in removing noise from electrical logging data using data science methodologies without effecting the true readings from formations?
2. Can our facies classification provide a more detailed lithological groupings within formations in comparison with mudlogs?
3. Which of the three wells exhibit better petrophysical parameters in terms of hydrocarbon accumulation?

III. DATA (MATERIALS)

In this research we are analysing three wells (15/9-F-11T2, 15/9-F-11A, 15/9-F-11B) from Volve Village dataset obtained from Azure platform. Data contains necessary Logging While Drilling (LWD) measurements with several scalar attributes in .las format.

Table 1 shows the logging measurements we will be using in this study and the corresponding description and units for each. The first two wells (15/9-F-11T2, 15/9-F-11A) contain all the necessary curves for analysis. However, 15/9-F-11B well has no sonic log measurements, which we will later try to predict using supervised machine learning.

TABLE I. LWD MEASUREMENTS

Curve Name	*Description*	*Units*
Gamma Ray (GR)	measures natural formation radioactivity	API
Resistivity (RMED, RDEEP, RT)	measures the rocks' ability to impede the flow of electric current	ohmm
Photoelectric effect (PEF)	measures electrons emission of a facies illuminated by light rays	barns/ electron
*Sonic Log (DTC, DTS)	measure of a how fast elastic seismic compressional and shear waves travel through the rock formation in transit time	ft/sec
Density (RHOB)	measures the bulk density of the associated rock formation	g/cm^3
Neutron Porosity (NPHI)	measures the amount of hydrogen atoms present in the pore spaces of the rock	v/v

* Not present in 15/9 F-11B

Additional data include mudlog files in .dlis format which contain lithological codes for cuttings of rock brought to the surface by the circulating drilling mud. This data will play a significant role in later comparison of our facies predictions. Petrophysical reports were also available in .pdf format, with all the evaluation parameters necessary for our calculations of porosity, permeability, shale volume and water saturation. Three formation tops (Heather, Hugin, Sleipner) near the reservoir section were acquired from geological reports.

From the machine learning point of view, we concatenate the available attributes along with calculated parameters and associate to each depth and well a feature vector to predict facies classifications using clustering algorithm.

IV. ANALYSIS

A. Data Preparation

The first step of the data preparation is the data exploration for quality control and to identify any missing values or noise. It is observed that missing values are only limited towards the foot of the borehole, since each tool making physical measurements are spaced along the instrument's length behind the drilling bit. Since none of the missing values were present in the Hugin reservoir, which we are evaluating, all the rows with missing values were dropped without concern.

One of the pre-processing steps was the application of the median filter, which replaces each entry with the median of neighbouring entries in an attempt to remove noise while preserving the relative magnitude of the curves.

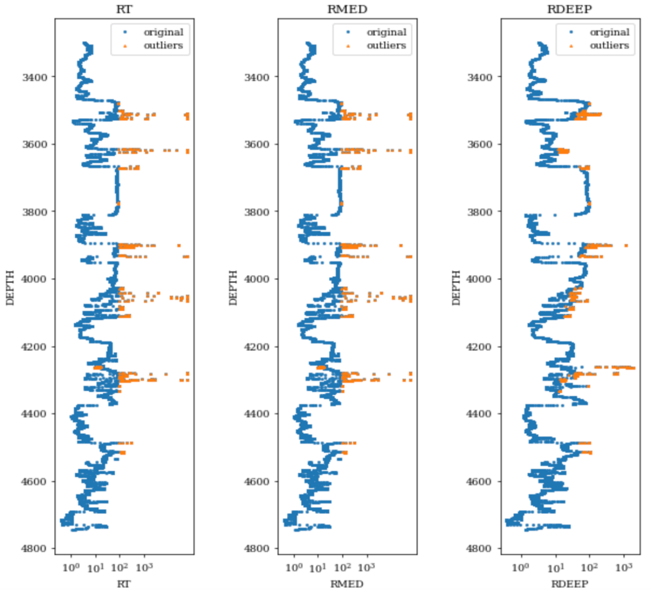


Figure 1. Noise removal from resistivity curves with polarization horns using Isolation Forest method.

Despite noise removal, resistivity logs in Hugin formation from 15/9-F-11B still showed outlier noise due to the affects of polarization horns, which is a charge build-up between two formation layers with different dielectric properties resulting in spike of high resistivity as the tool crosses the bed boundary. We managed to remove these novelties in resistivity data using Isolation Forest outlier removal method with a threshold of 0.6 (figure 1).

Additionally, smoothing was applied to raw measurements, to account for thin formation layers. Higher smoothing window (equivalent to 4

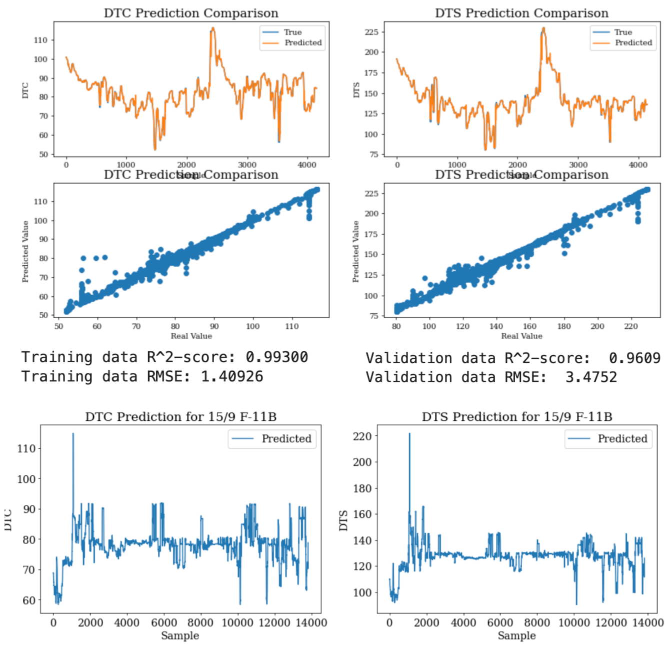


Figure 2. Sonic log prediction results for well 15/9 F-11B using Random Forest.

meters) was applied to 15/9-F-11B, since it consisted of more thin layers than the other two wells.

As an added step of the pre-processing, we attempted to generate a synthetic sonic log measurement [2]. Using existing sonic log data from other two wells, we trained and validated our model using Random Forest in order to predict and account for the missing DT and DTS curves in 15/9-F-11B (figure 2). Last but not least, we loaded mudlog cutting description data in .dlis format, which later was merged with the raw measurements.

B. Data Derivation

The data derivation process consists of calculation of several petrophysical properties such as Shale Volume (VSH), Total Porosity (PHIT), Horizontal Permeability (KLOGH), and Water Saturation (SW), using evaluation parameters described in the petrophysical report provided by Equinor ("Hugin and Skagerrak Formation, Petrophysical Evaluation", November 2006).

PHIT (ϕ_T) is defined as the ratio of the entire pore space in the rock to its bulk volume. It is derived from Bulk Density (RHOB) and Neutron log (NPHI) measurements, both of which have been corrected for varying mud filtrate invasion.

$$\varphi_D = \frac{(\rho_{matrix} - RHOB)}{(\rho_{matrix} - \rho_{fluid})}$$

$$\varphi_T = \varphi_D + A \times (NPHI - \varphi_D) + B$$

where,

- ρ_{matrix} is the matrix-density
- $RHOB$ is the measured bulk-density
- ρ_{fluid} is the pore fluid-density
- φ_D is density porosity
- A and B are regression coefficients.
- $NPHI$ is neutron log in limestone units

Shale Volume was derived from linear gamma ray relationship.

$$VSH = \frac{(GR - GR_{min})}{(GR_{max} - GR_{min})}$$

where,

- GR is the measured gamma ray
- GR_{min} is gamma ray reading in clean sands
- GR_{max} is gamma ray reading in shales

The horizontal log permeability (KLOGH) is derived from the following equations based on existing multivariable regression analysis between log porosity and shale volume against overburden corrected core permeability:

- Hugin: $KLOGH = 10^{(2+8\times PHIT-9\times VSH)}$
- Sleipner: $KLOGH = 10^{(-3+32\times PHIT-2\times VSH)}$

Water saturation is calculated using Archie equation, giving a total water saturation [3].

$$Sw_T = \left[\frac{a \times R_w}{\varphi_T{}^m \times R_T}\right]^{\frac{1}{n}}$$

where,

- a is Archie (tortuosity) factor
- R_w is resistivity of formation water
- φ_T is Total Porosity
- m is cementation exponent
- R_T is true resistivity measurement from well logs
- a is saturation exponent

Water resistivity of formation at formation temperature was calculated using below equation.

$$R_w = R_w(@TR_w) \times (TR_w + KT_1)/(F_T + KT_1)$$

where,

- $R_w(@TR_w)$ is R_w measured at temperature
- TR_w is temperature at which R_w was measured
- F_T is Formation Temperature ($111°C$)
- KT_1 = 21.5 (for Metric units)

Bulk Volume of Water is the proportion of the rock that is estimated to be formation water and is found by multiplying fractional porosity by the fractional water saturation.

$$BVW = S_w \times \phi_T$$

Sand Flag was derived by using PHIT > 0.10 and VSH < 0.50 and KLOGH > 20 conditions.

C. *Construction of Model*

Many different and complex supervised classification techniques have been developed through the years [4]. However, considering that we will be using limited amount of data, the study relies on a simpler classification technique. K-Means is one of the simplest unsupervised learning algorithms that can solve problems around facies predictions when no training data is available [5].

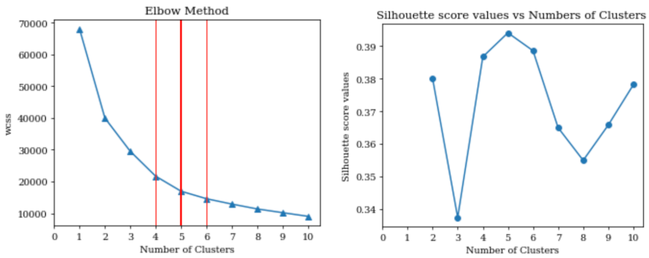


Figure 3. Elbow method and Silhouette score values vs. Number of Clusters

The procedure follows a simpler and easy way to classify a given data set through a certain number of clusters. RHOB, NPHI log measurements along with KLOGH, PHIT and VSH calculated properties were scaled to be used in the model. Elbow method and silhouette scores were employed to select the right number of clusters [6]. In figure 3, we can see that the inertia (sum of the squared distances to the nearest cluster centre) decreases as we increase the number of clusters. There is no clear defined break in the Elbow method, however, we can see that the Silhouette score values change drastically in about 5 clusters onwards. Hence, total of 5 clusters corresponding to different lithofacies were generated and plotted against each feature for visual analysis of how the clusters vary (figure 4).

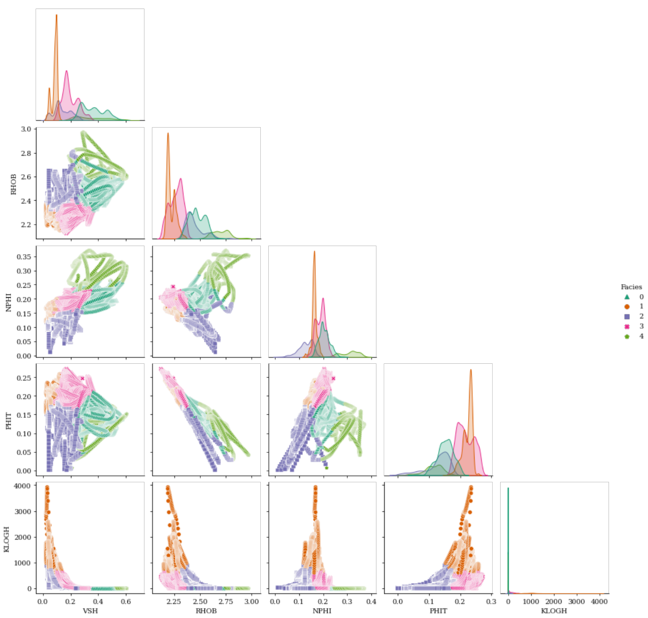


Figure 4. Predicted facies pair-plot w.r.t. features.

D. *Validation of Results*

As seen in the figure 4, the petrophysical features such as permeability and porosity calculated from well logs' response have high similarity for the same facies type and data points with similar traits are assigned to the same group: Facies 1 and 3 having the best rock type (we will rename them to A and B), Facies 2, 0 and 4 (C, D, E) being most shaly and corresponding to a poorer lithology.

The only labelled lithology available to us is from mudlogs, which tend to give a very general lithological information about the formations. However, we can still plot our predicted facies in log plots in relation to labelled cuttings' descriptions to see how well our unsupervised method perform with logging data. As seen in the figure 5, 6, 7, corresponding to each well, the facies predicted through K-Means clustering correlate well with lower-level lithology codes obtained from mudlogs, which only provide basic overview of the lithologies encountered, while clustered facies give a more detailed litho-classifications.

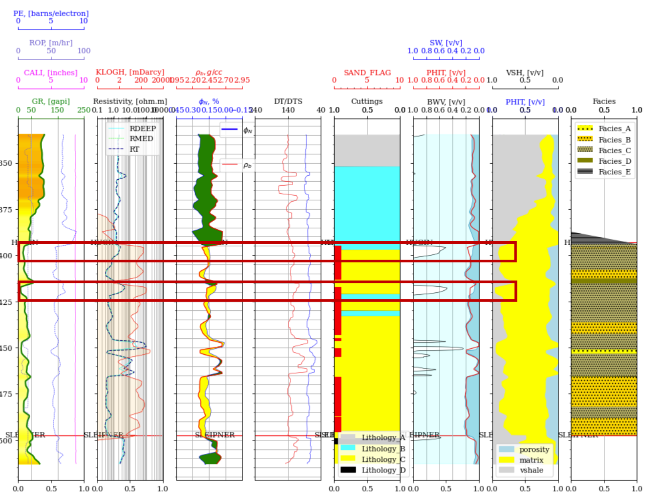


Figure 5. Composite Log plot for 15/9-11T2 including petrophysical and facies classification results.

V. Findings, Reflection and Further Work

Answering to the first analytical question of this study, we were able to find out that by applying median filter we can remove most of the noise in the log curves without effecting the relative magnitude of the curves. For special cases, such as polarization horn effects in resistivity measurements, employing Isolation Forest outlier detection method produces best results without skewing the actual rock measurements. As part of the future work, we can experiment with other machine learning algorithms, such as One Class SVM, Local Outlier Factor and Elliptic Envelope in removing outliers. Additionally, we employed machine learning algorithm (Random Forest) in predicting curves that weren't recorded either due to financial implications or borehole conditions. All these advanced statistical and machine learning techniques can help reduce the uncertainties in hydrocarbon predictions which can bring value in terms of time and money to the company.

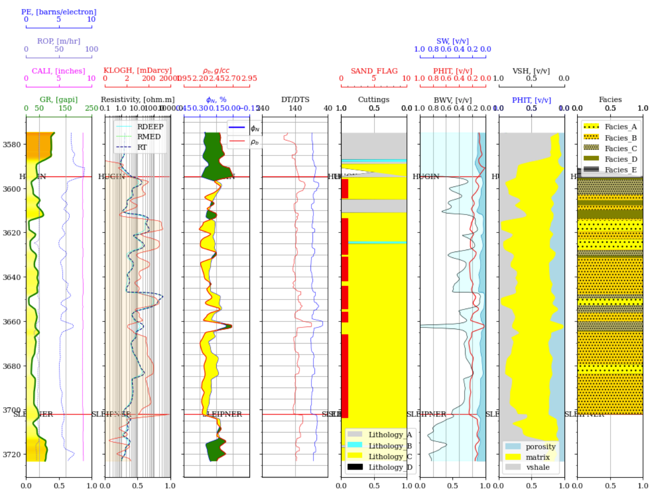


Figure 6. Composite Log plot for 15/9-11A including petrophysical and facies classification results.

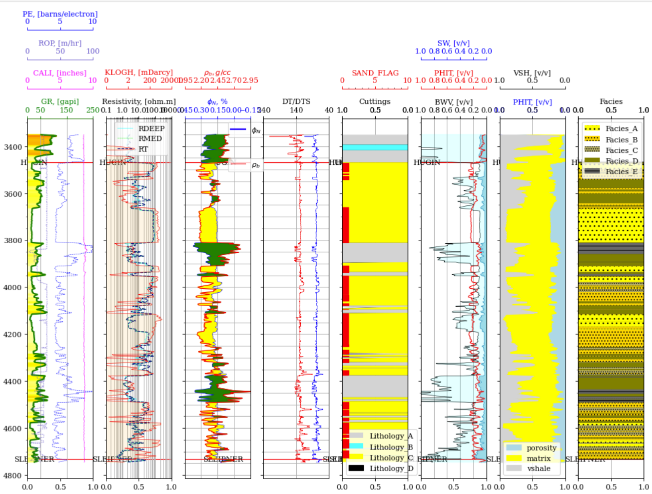
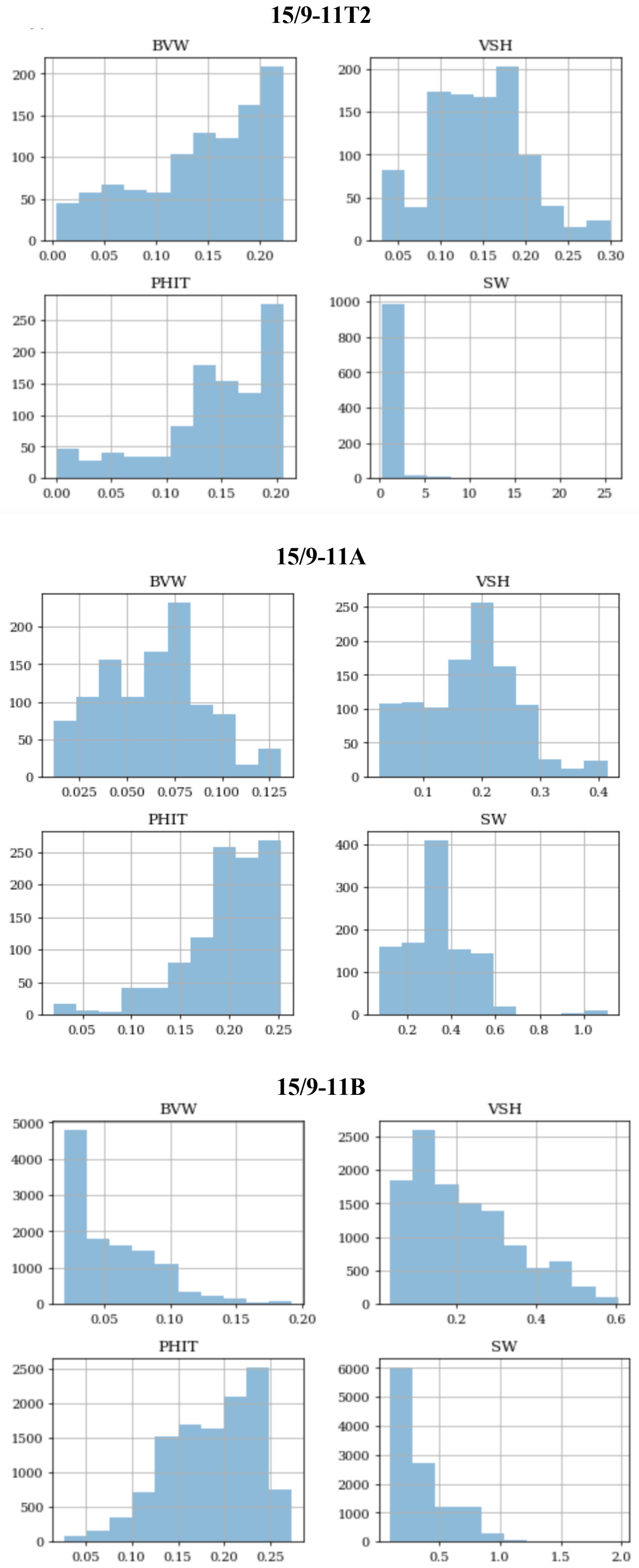


Figure 7. Composite Log plot for 15/9-11B including petrophysical and facies classification results.




```
==================== ==================== ====================
====== F-11 T2 ===== ====== F-11 A ====== ====== F-11 B ======
ZONE: HUGIN FORMATION ZONE: HUGIN FORMATION ZONE: HUGIN FORMATION
 4393.0 m- 4497.4 m    3594.6 m- 3702 m      3467.5 m- 4732 m
Total: 104 m          Total: 107 m          Total: 1264 m
Mean values:          Mean values:          Mean values:

BVW     0.144200      BVW     0.063605      BVW     0.055401
VSH     0.143958      VSH     0.180910      VSH     0.215728
PHIT    0.143083      PHIT    0.196091      PHIT    0.186442
SW      1.144632      SW      0.337271      SW      0.341774
```

Figure 8. Hugin formation petrophysical parameters summary.

As to the second question we were able to answer that in situations where we don't have actual lithological data a method for classifying facies from indirect electrical measurements from well logs is essential. Looking at well 15/9-F-11B (figure 7), we have 5 separate facies/groups displayed and we can see that these mostly tie up with the changes in the logging measurements. For example the decrease in Gamma Ray from around 3650m to around 3800m ties in nicely with the yellow Facies-A grouping. In the Mudlogs however, this section has also been highlighted as being in the same cluster, however, there is no variation with the section above 3650m. As mudlog cutting descriptions are just the rock mixtures brought to the surface by the circulating drilling mud, we will not be able to capture the same degree of variations, which proves the importance of facies classification through clustering method using log measurements. As part of the future work, we can try using other clustering techniques such as Gaussian Mixture Modelling (GMM). Furthermore, we can try including all other wells from Volve field, incorporating exploration wells with core data, to extend the scope of our research.

As to the third question, looking at the overall picture using the log plots and the Hugin formation summaries, we can see that well 15/9-11B does provide a much better reservoir properties and better rock quality based on the facies classification (figure 7 and figure 8). This well also exhibits the thickest net sand in Hugin formation. In 15/9-F-11-T2 on the other hand, Hugin Formation seems water filled, except of two pockets of residual oil that seem to be structurally trapped (figure 5, red boxes indicating residual oil). This well also exhibits hardly any Facies A type facies which is best quality rock. 15/9-F11-A shows similar thickness to 15/9-F11-T2 but with a lot more Facies A rocks, and Oil Down To (ODT) ~3700 m measured depth (MD).

All of our observations of three wells thus far agree with the petrophysical interpretations made in Equinor reports. There are 8 perforated intervals in 15/9-F-11-B, and as an additional future work, we could integrate production data with this study to extend the scope of this research.

References

[1] Bestagini, P., Lipari, V. and Tubaro, S. (2017) "A machine learning approach to facies classification using well logs," in SEG Technical Program Expanded Abstracts 2017. Society of Exploration Geophysicists.

[2] Yu, Y. *et al.* (2021) "Synthetic sonic log generation with machine learning: A contest summary from five methods," Petrophysics – The SPWLA Journal of Formation Evaluation and Reservoir Description, 62(4), pp. 393–406.

[3] Archie, G. E., (1952) "Classification of carbonate reservoir rocks and petrophysical considerations". AAPG Bulletin, vol 36, no.2, pp 218-298.

[4] Bishop, C. M., (2006), "Pattern recognition and machine learning". Springer-Verlag New York, Inc.

[5] Macqueen, J. B. (1967) "Some Methods for classification and Analysis of Multivariate Observations". Proceedings of 5th Berkeyeley Symposium on Mathematical Statistics and Probability. 1. University of California Press. pp. 281-297

[6] Robert L. T., (1953) "Who Belongs in the Family?". Psychometrika. 18(4). pp. 267-276

Word Counts

Abstract - 104

Introduction - 201

Analytical Questions - 153

Data (Materials) - 306

Analysis - 1007

Findings, Reflection and Further work - 512