ONLINE PRESENTATION
24 - 25 OCTOBER 2020
PROFESSIONAL TECHNICAL PAPER

# Machine Learning Application on Formation Testing Result Status Prediction – A Case Study

Ratna Dewanda[*1], Aulia Desiani Carolina[1,2], Jhonny[1,3], Hengky Ng[1,4], Rinaldi Sudarwoto[1,5], Imam Nugroho[2]

[1]Schlumberger,
[2]Geopysical Engineering, Universitas Pertamina
* Email: rdewanda@slb.com

**Abstract.** Data-driven recommendations and decisions play an important role in today's challenging oil and gas industries. Incorporating automation offers the ability to analyze bigger, more complex data and deliver fast results – even on a very large scale. Machine learning is one of the branches of Artificial Intelligence which can automate the analytical model building by learning from data, identify patterns and make decisions with minimal human intervention. On this study, machine learning method was applied on a set of data from one field to predict the pressure test or pretest status of wireline formation testing result for a given depth with a model built based on the well logs. First, rock quality classification was performed using an unsupervised approach. Next, using the previous step output as a constraint, Neural Networks method was applied to predict the status of wireline formation testing pretest result. The result of pretest prediction was then validated with actual pretest result. This study is using the dataset of four wells as "train" data and predicting the result status of one "target" well. The accuracy of prediction of the "train' dataset was above 80% and the accuracy for the "target" well was 100%. This difference might be explained by the wide range of data that was incorporated as "train" dataset wells which able to build a robust model to predict the "target" well accurately. This study shows the application of machine learning on good data set will leverage the value of data providing different perspective from the conventional decision-making process.

**Keyword:** Machine learning; python; unsupervised; classification; Neural Network; prediction; formation testing; pretest

## 1 Introduction

Basic Wireline Formation Tester (WFT) data, such as formation pressure and drawdown mobility, play important roles for oil and gas (O&G) operators in determining the location of oil/gas sweet spots in the particular well, the connectivity between reservoir layers, selecting the best sampling points, and the assessment of the reservoir's lifting capability to the surface. Therefore, obtaining a valid pressure test (pretest) status is an important goal in WFT logging. Unfortunately, the complexity of the reservoir, the quality of the reservoir rock, and the challenging logging environment, often leads the WFT logging operation to unsuccessful pretest results, such as supercharged test, tight or dry test, and lost seal test. These unsuccessful pretests in the end will reduce the pretest's success ratio, which determines the efficiency and the effectiveness of the WFT logging operation. Therefore, optimizing the WFT logging program by performing the pretest point candidate selection is crucial for both operators and service providers in order to increase the likelihood of getting valid test status from the reservoir during the WFT pretest. Conventionally, this pressure point candidate selection is performed by a group of geoscientists, including geologist, petrophysicist, and reservoir engineer, which manually evaluate open-hole logs data, consisted of triple combo log, acoustic log, image log, and nuclear-magnetics log, to select the best pressure point candidate. However, since this approach is done manually, the process sometime can be cumbersome, and the result may be inconsistent from one interpreter to another which in the end would consume a significant time on decision making process.

In the dawn of digital era, the application of Artificial Intelligence (AI) revolutionize the existing conventional workflows in the industries, including O&G industry. The application of digital computation has tremendous potential in enabling the O&G users to make faster consistent reliable decision, including to optimize the WFT pretest point selection using AI algorithm. Machine Learning (ML) is a subdivision of AI application, which generally uses a set of data to learn the pattern and the relationship thereof, using abundance of statistics and mathematical equations written in computer algorithms in order to predict a desirable outcome. ML workflow incorporating automation offers the ability to analyze bigger, more complex data and deliver fast results, even on a very large scale with high accuracy of prediction. It is more widely used nowadays including in oilfield industries as the data driven decision becomes a necessity. On this study, the applications of machine learning methods to predict the pressure test or pretest status of wireline formation testing result are explored.

The data sets used on this study are from G-14 field, Netherland. Data screening was done based on the quality and availability of the data coming from the same field. In this study, five well data sets were used. From those 5 (five) wells, four wells served as "train" data which were used to train and validate the ML model and 1 (one) well served as "target" well to which the learned model be applied. Those five wells have open-hole logs, including but not limited to, Gamma Ray (GR), Resistivity (RES), Density (DEN), Neutron (NEU), and WFT data of pretest status, formation pressure, and drawdown mobility. In this study, the WFT pretest status was simplified into "Good" test and "Bad" test based on the acquired formation pressure and drawdown mobility measurements. Some log data sets were normalized and cut to the relevant intervals as deemed necessary for the study.

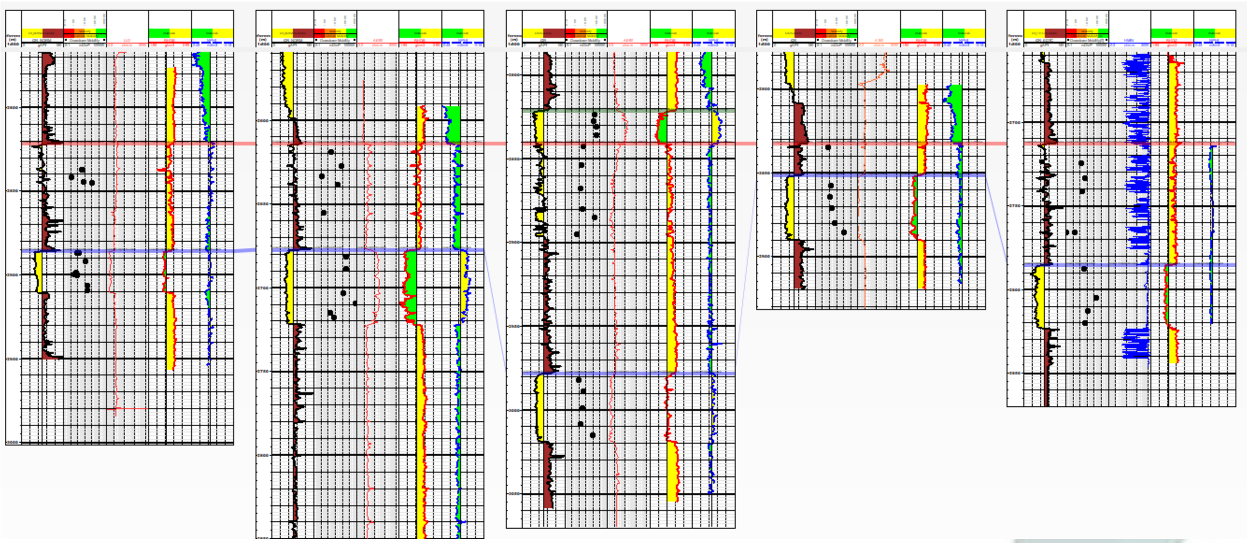

*Figure 1 – Well logs and WFT result from a set of wells in the G14 Field, Netherland*

## 2 Methodology

There are two ML algorithm approaches that will be used to predict the WFT pretest status result. The first ML algorithm approach (ML-1) was a-built-from-scratch workflow using python script, while the second ML algorithm approach (ML-2) was based on a specific module inside Techlog™ software – a Schlumberger technology platform application for well log analysis. Both ML algorithm approaches were using Unsupervised ML methods and Supervised ML methods.

The Unsupervised Machine Learning (UML) method was conducted to generate the rock type function using a specific rock typing method for each ML algorithm approach. Meanwhile, the Supervised Machine Learning (SML) method was performed to predict the WFT pretest status results using a defined neural network workflow for each ML algorithm approach incorporating the input from open-hole logs data and rock typing function generated from UML methods.

The simplified methodology of both ML algorithm approached workflow is illustrated in the Figure 2 below. Each of the UML and SML methods for both ML algorithm approaches will be described in the following sections.

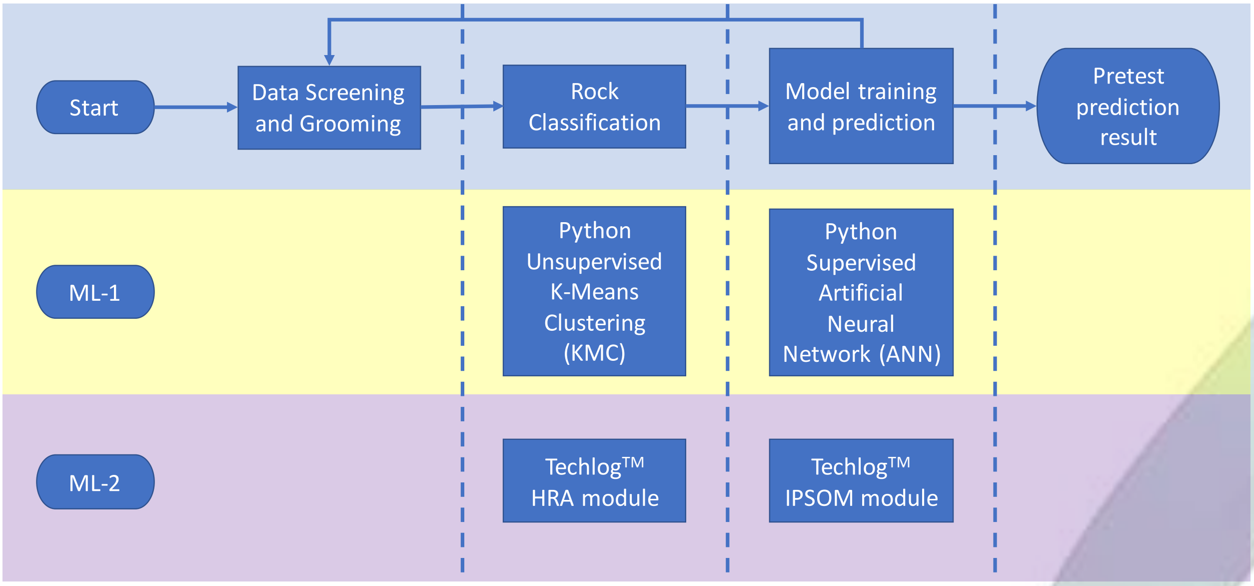


*Figure 2 – Simplified Machine Learning (ML) algorithm approach workflow*

## 2.1 ML-1 algorithm approach

The First ML algorithm approach (ML-1) was performed using both UML and SML methods, whereby the UML method was based on K-Means Clustering (KMC) for rock typing and the SML method was based on the neural network algorithm for predicting the WFT pretest status results.

### 2.1.1 K-Means Clustering (KMC)

K-Means Clustering (KMC) is a method of vector quantization which used for classifying observation or measurements into K clusters where each observation belongs to the cluster with the nearest mean or center or centroid and minimized within cluster variances. To process the learning data, KMC starts with first group randomly selecting the centroids, which later used as the beginning points for every cluster and then perform iteration computation to optimize the location of the centroids. In this study, the chosen KMC algorithm was compared with other methods for the rock typing in order to obtain better rock typing classification. There are four rock types classification for the 5 (five) wells being generated using this method, such as “Very Good” rock, “Good” rock, “Intermediate” rock, and “Non-Reservoir” rock.

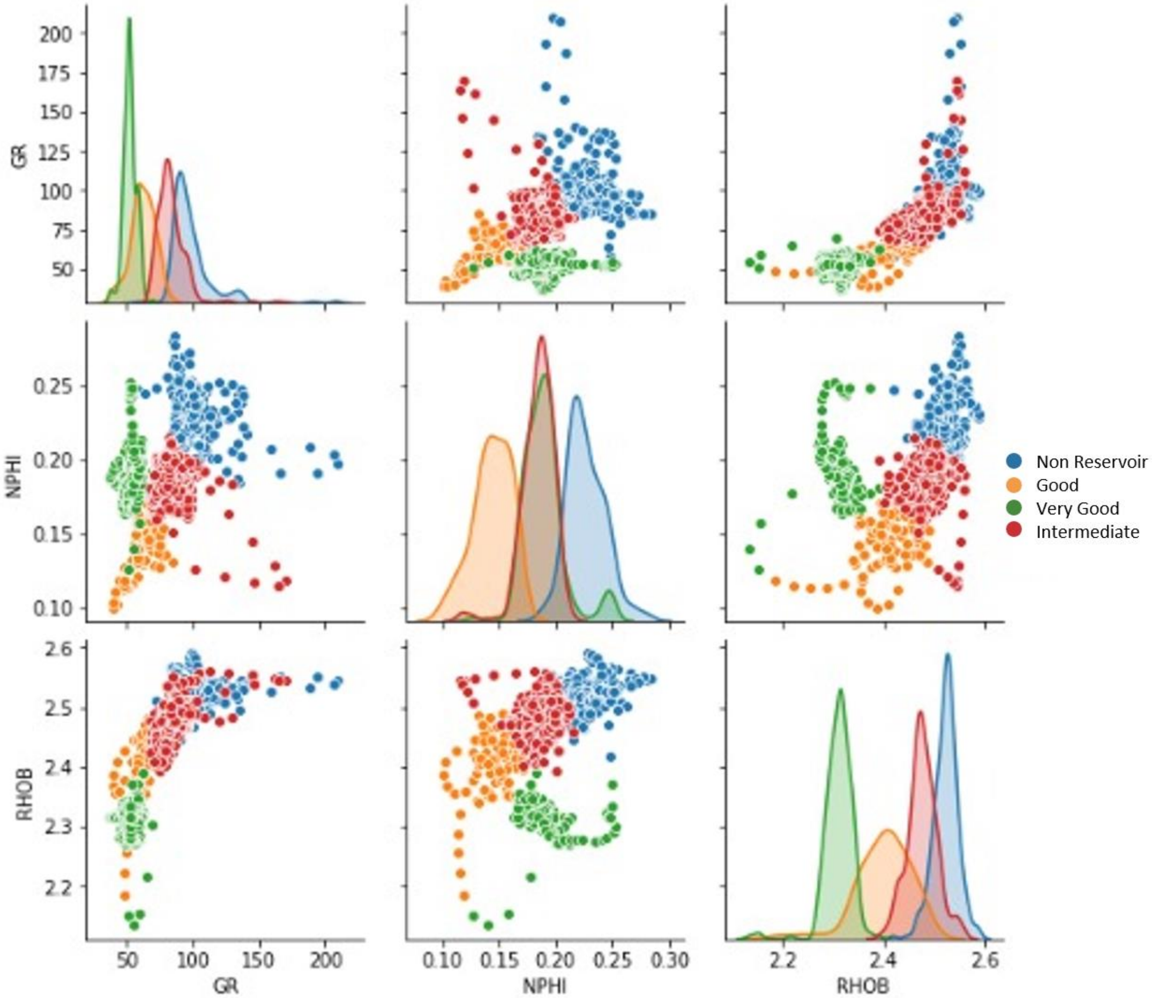


*Figure 3 – Gamma Ray – Neutron Porosity – Bulk Density Rock Type Classification in Well G14-02*

### 2.1.2 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is one of computational model parallels that copy physiology from human brain system (Mitchell, 1997). It will send the weight on each input to each node, then it will calculate and will be compared with threshold within activation function. The calculation will continue to next layer if the parameter is above the threshold and will be back again to make the accuracy comes near to the best accuracy.

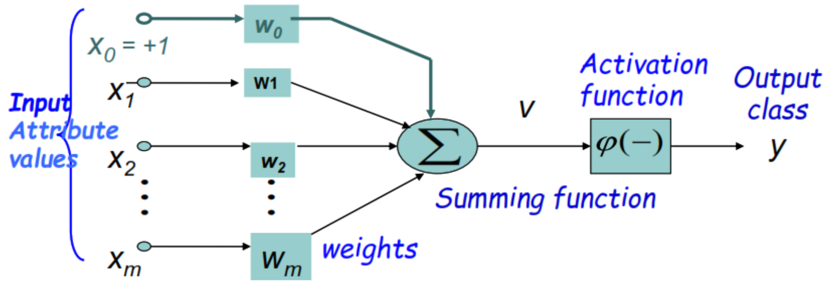


*Figure 4 – Artificial Neural Network (ANN) Illustration (Kumar, 2003)*

ANN contains several parameters such as: number of layers, activation per layers, number of nodes, how many iterations, and calculation of error. In this study, the existing input data and previous additional data from KMC were used as input for ANN computation. The result from ANN computation will contain of two answers, which are “Good” and “Bad” tests.

### 2.2 ML-2 algorithm approach

The Second ML algorithm approach (ML-2) was also performed using the UML and SML methods but from the specific module built inside Techlog™ software, whereby the UML method was performed using Heterogeneous Rock Analysis (HRA) for rock typing, while the SML method was performed using Indexation Probability and Self Organized Map (IPSOM) for predicting the WFT pretest status results.

#### 2.2.1 Heterogeneous Rock Analysis (HRA)

Heterogeneous Rock Analysis (HRA) is a log-based rock classification method developed for the integration of core data and correlation of core data to logs in unconventional reservoirs. HRA defines rock classes based on their fundamental attributes of texture and composition as discriminated by log inputs and used to discriminate the material properties of the rocks.

Algorithmically, HRA identifies consistent data structures, defined initially by unsupervised pattern recognition of the input data channels, for example, open-hole log data. The unsupervised classification is thus predicted on the structure of the data variance, and not on pre-conceived ideas of what these classes should represent. The resulting patterns have a unique meaning in texture and composition space. Consequently, the rock classes become uniquely recognizable (Doveton, 1994, Handwerger, 2011) through supervised pattern recognition on subsequent wells, or even zones external to the modeled interval within a single well.

By extension, the properties correlated to the rock classes are then transferrable to non-cored wells or sections via the classification, even if not directly solvable to through other deterministic or inversion models of the logs themselves. HRA applies to any type of multivariate data, such as log, core, mud logging, seismic data, etc. By using the HRA, it is possible to link and integrate between these measurements. In this case, the open-hole log data was used as reference for measurement and scale. Due to the open-hole logs are regionally prevalent, they are in a useful scale to solve field problems, they have numerous data channels that tend not to be mathematically dependent on each other or serve as independent measurements of the same rock and O&G operators are used to them.

In this study, Gamma Ray (GR), Bulk Density (DEN) and Neutron Porosity (NEU) used for rock typing using HRA module. The HRA module first runs Principal Components Analysis (PCA) to transform the input data onto independent axes which front-load the variance and ensure that the data used in the clustering are functionally independent. After the PCA analysis, the principal components are used in a KMC algorithm to create the HRA classification. Several plots are automatically generated to help in performing quality control the clustering process and choose the number of clusters required to best classify the rock. Similar to ML-1 algorithm approach, there are four rock type choose as the output from the classification, which are Very Good" rock, "Good" rock, "Intermediate" rock, and "Non-Reservoir" rock. Those 4 (four) rock types have good represent the log respond and consistent for all 5 (five) wells.

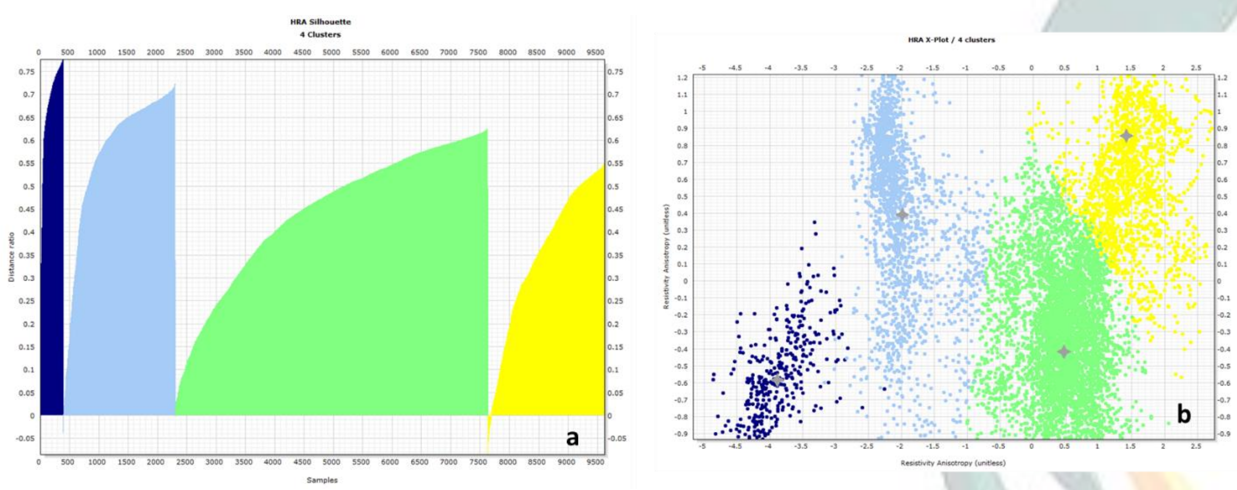


*Figure 5 – Silhouette and Centroid distribution of the rock types for all 5 (five) wells*

2.2.2 Indexation Probability and Self Organized Map (IPSOM)

Indexation Probability and Self Organized Map (IPSOM) is an Artificial Neural Network (ANN) which using single or multi-variable data input to identify patterns/groups in data using the principles of Self-Organizing Map (SOM), Indexing and Probability. The SOM is an ANN which is trained and represented in a 2-dimensional view. This was first described by T. Kohonen and is therefore sometimes referred to as the Kohonen map. The learning algorithm of SOM is an unsupervised learning algorithm where the data is transformed into nodes that represent the distribution of the data when the learning algorithm launch.

Once the learning is complete, and the data is upscaled to a user-defined set (how many nodes used to represent the data). Then, it will use Indexing to assign a classification to each node. There are two types of indexing available unsupervised classification and supervised classification. The unsupervised indexation assigns a classification to each node on the Kohonen map. User decides how many classes to output from this unsupervised indexation. The supervised indexation works when there is indexation or classification data available as the index input. The use of this input is to calibrate the Kohonen map. In order to check the confidence level of the result, there is a probability curve that can be used to understand the uncertainty involved with the classification, which allow user to know probability of each depth of the classification.

In this study, IPSOM was used to predict the WFT pretest status result for the “target” well, G-14-03, which assume no WFT data available on that well. Together with the Rock Classification (output from HRA) and the WFT pretest status from the 4 (four) other “train” wells, the open-hole logs, such as GR, RES, DEN, NEU, including caliper log (CAL), and slowness (DT) were used as the input data.

For indexation method, supervised with minimal distance method was selected, where the minimal indexation method reviewed the location of every data point inside the node and the distance of every data point is calculated from the node value. The depth point which is closest to the node value is taken, and the indexation value at this depth is assigned to the node.

Once the method is launched, ranking variables are generated to rank the correlation between input with the output. This allows to remove the inputs that are considered least correlated to the output and improve the model. The result was compared with the actual pretest status and the process was relaunched until overall accuracy level met the predefined expectation (>80%).

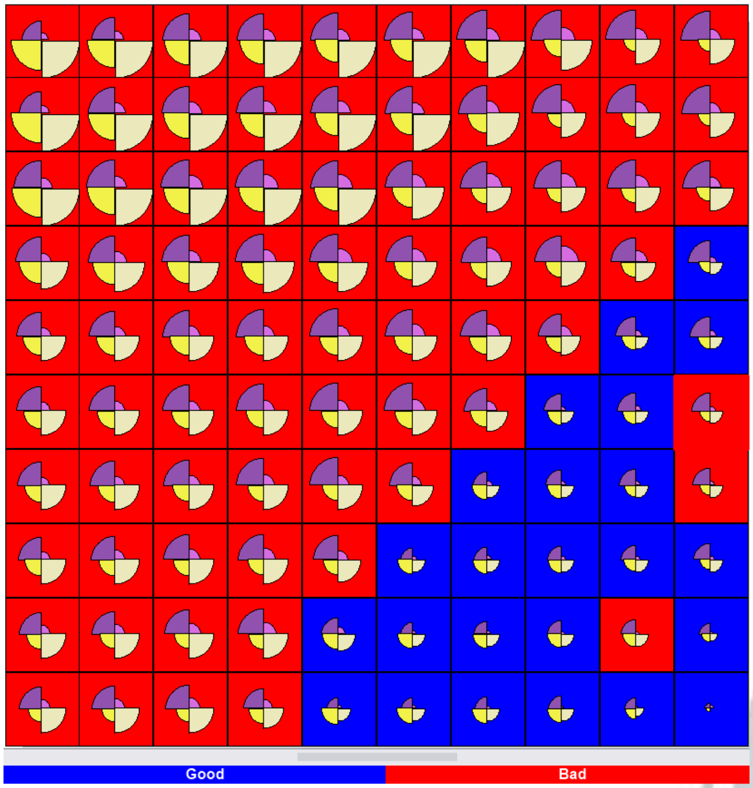


*Figure 6 – Self Organized Map or Kohonen map of the study area*

## 3 Result and Discussion

### 3.1 Rock Typing Classification (Clustering)

As previously mentioned, there are two ML algorithm approaches carried out in this study. Both approaches are using the same terminology for the 4 (four) defined rock types, which are “Very Good”, “Good”, “Intermediate”, and “Non-Reservoir” rock types. Based on the observed output data, both KMC and HRA from each UML method from ML-1 and ML-2, respectively, output similar results for the main reservoir zone. The result from the ML-1 algorithm approach is comparable with the result from ML-2 as shown in Figure 7.

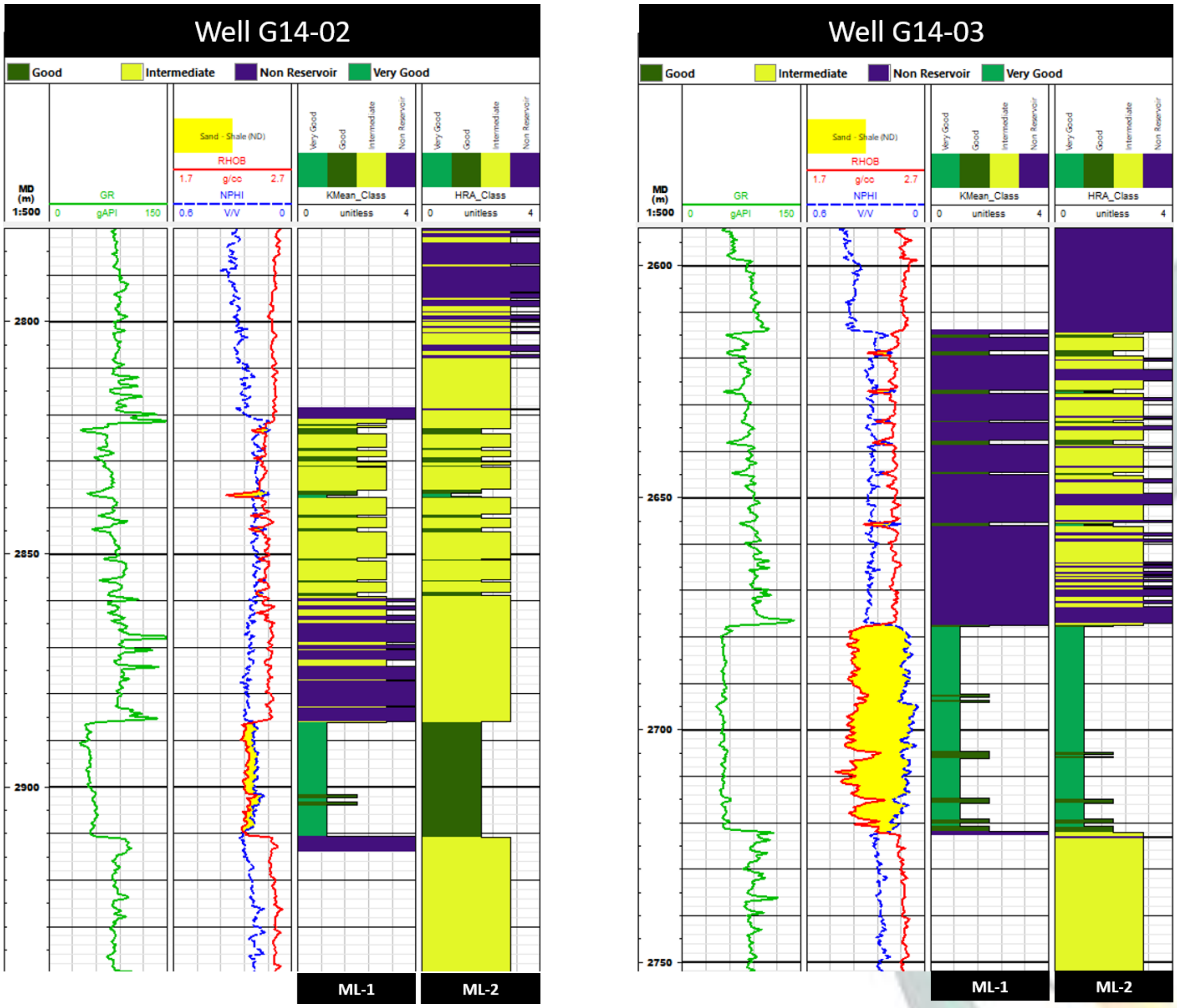


*Figure 7 – Rock Type Classification on Well G14-02 and G14-03 based on ML-1 and ML-2 approaches*

Figure 7 shows the clustering results where the rock types were divided into 4 (four) classification based on the centroid of each cluster. This rock types classification used Gamma Ray, Resistivity, and Density-Neutron logs as input data. Those open-hole logs were chosen due to its consistency and availability in all "train" and "target" wells. Some differences occurred over the rock types classification result from ML-1 and ML-2 approaches were expected. In the ML-1, all the inputs are used directly in the KMC method, whereas in the ML-2, all the inputs were firstly transformed into Principle Component Axes (PCA), followed by clustering the data based on the KMC method. Apart of the difference in algorithms, the number of samples of input data also plays important role in affecting the results. The ML-1 input data was cut into the target interval, hence was shorter compared to the ML-2. However, both results are still comparable and consistently identify the "Very good" and "Good" rock against the "Intermediate" and "Non-reservoir" rock. These can be used for the next step in predicting the WFT pretest status.

There are 2 (two) main differences of clustering between the ML-1 approach (KMC only) versus the ML-2 approach (HRA with PCA plus KMC). First, the ML-2 approach has Silhouette display (Figure 5) that can be used easily to check the optimum number of clusters to define the number of rock types in formation. Second, the ML-2 approach has only one cross-plot display (Figure 5) to validate the clustering result (the x-axis and y-axis in the cross-plot are the main Principal Component – PC1 and PC2). On the other hand, the ML-1 approach has more cross-plot displays (Figure 3), as shown in this study, to validate of the clustering result.

3.2 WFT Pretest Status Prediction

The next step is to predict the WFT pretest status result by utilizing both ML approaches. For this step the accuracy criterion was set at more than 80% to qualify the fitness of the model. In the first approach, ML-1, the WFT pretest status prediction for the "target" well data (G14-03) was done by incorporating the ANN using GR, RES, DEN, NEU, CAL, DT, Rock Types Classification, and WFT data from the "train" well data (G14-02, G14-04, G14-05, and G14-06). For the ML-1 approach, the accuracy WFT pretest status prediction result from the "train" well data was around 86.7% and 100% for the "target" well data. The accuracy from this approach helps to build the confident level for the WFT pretest status prediction.

Similar to the ML-1 approach, the ML-2 approach predicts the WFT pretest status for the "target" well data (G14-03) using GR, DEN, NEU, Rock Types Classification, and WFT data from the "train" well data (G14-02, G14-04, G14-05, and G14-06). The difference is, for the ML-2 approach, IPSOM module in Techlog™ was used to predict the WFT pretest status instead of conventional ANN. As a result, for the ML-2 approach, the accuracy WFT pretest status prediction result from the "train" well data was around 83.2% and 100% for the "target" well data, with the average probability mean of 0.999 and the average standard deviation of 0.096. IPSOM automatically generated the probability curve that was used to quality control the result and build in the confident level for the WFT pretest status prediction. The summary of the "train" and "target" WFT pretest status prediction results for both ML algorithms are shown in the Table 1 below.

*Table 1 – Summary of prediction accuracy, probabilistic mean, and standard deviation for the "train" well dataset and "target" well dataset*

| Well | | ML-1 | ML-2 | | |
|---|---|---|---|---|---|
| **Name** | **Dataset** | **Accuracy** | **Accuracy** | **Probabilistic Mean** | **Standard Deviation** |
| G14-02 | Train | 95.7% | 75.0% | 0.999 | 0.106 |
| G14-04 | Train | 86.7% | 93.3% | 0.999 | 0.125 |
| G14-05 | Train | 87.5% | 87.5% | 0.999 | 0.063 |
| G14-06 | Train | 76.9% | 76.9% | 0.999 | 0.089 |
| **Overall Train Data** | | **86.7%** | **83.2%** | **0.999** | **0.096** |
| G14-03 | Target | 100.0% | 100.0% | 0.999 | 0.063 |

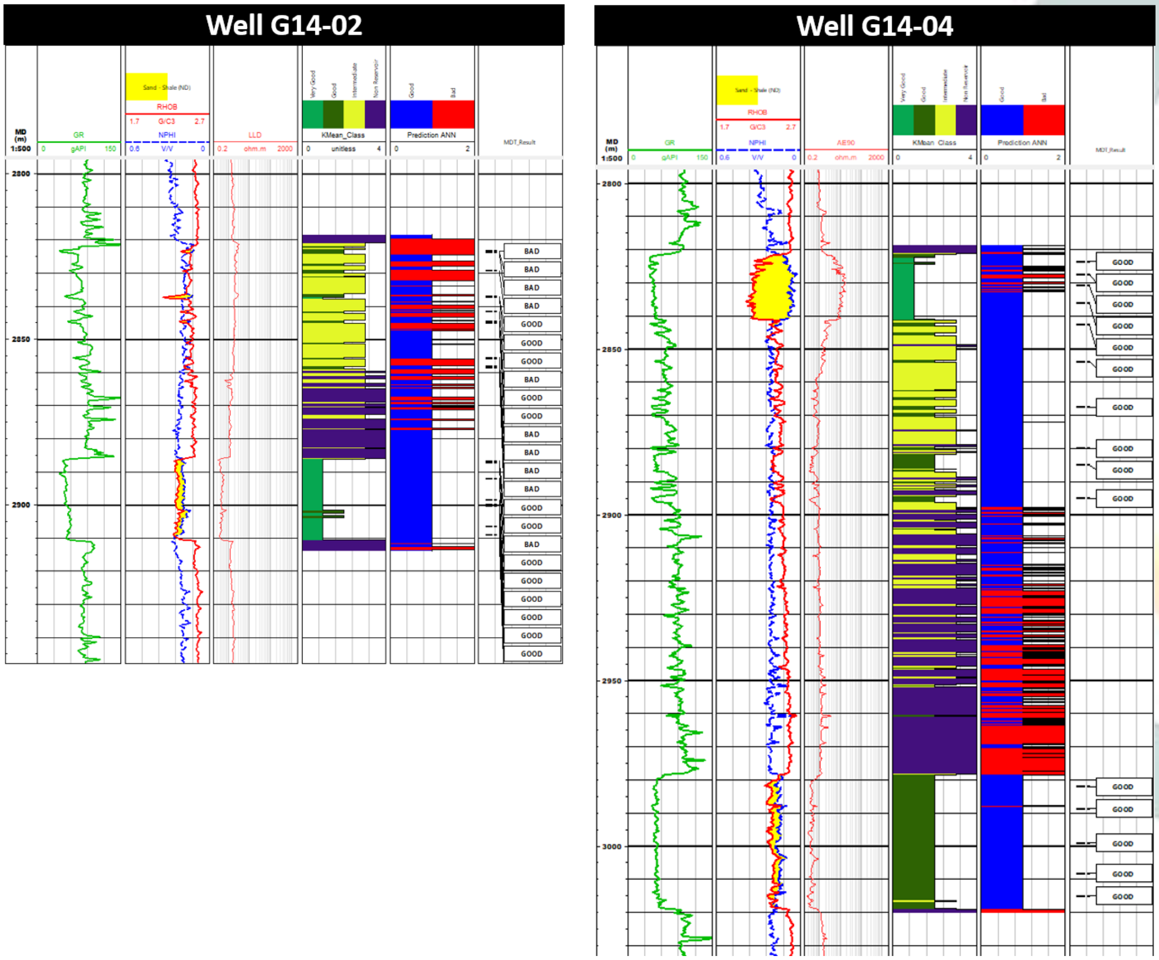


*Figure 8 Formation testing result using ANN for Well G14-02 and G14-04 as training dan testing wells*

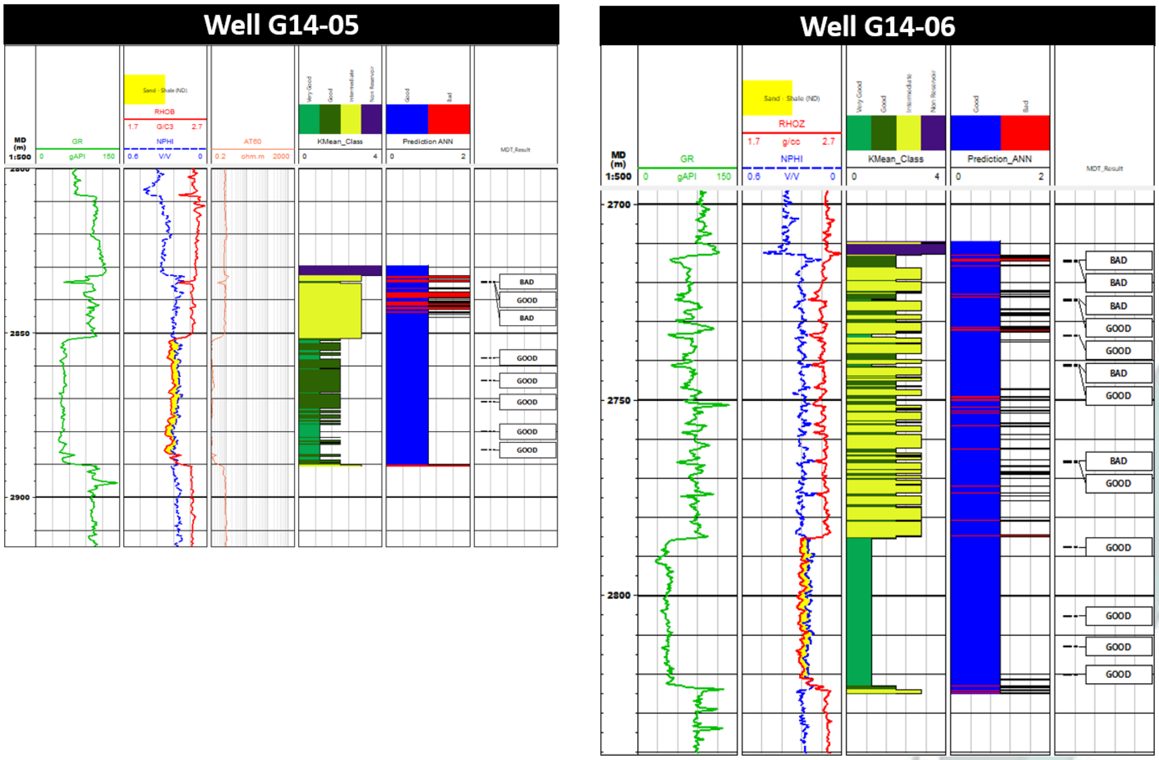


*Figure 9 Formation testing result using ANN for well G14-5 and G14-06 as training wells*

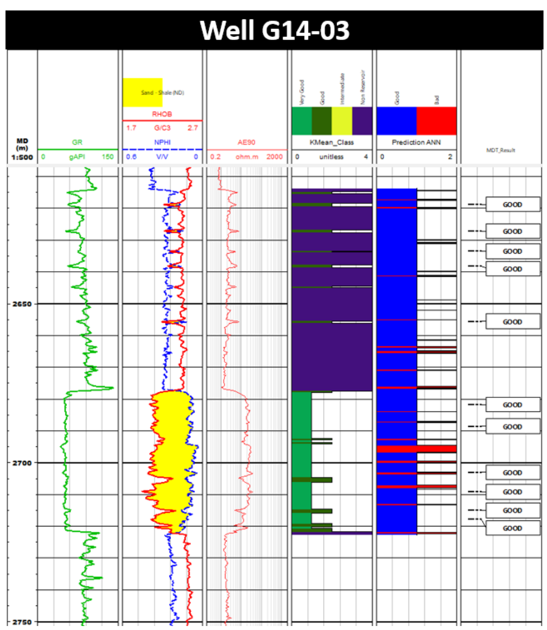


*Figure 10 Formation testing result using ANN for Well G14-03 as target well*

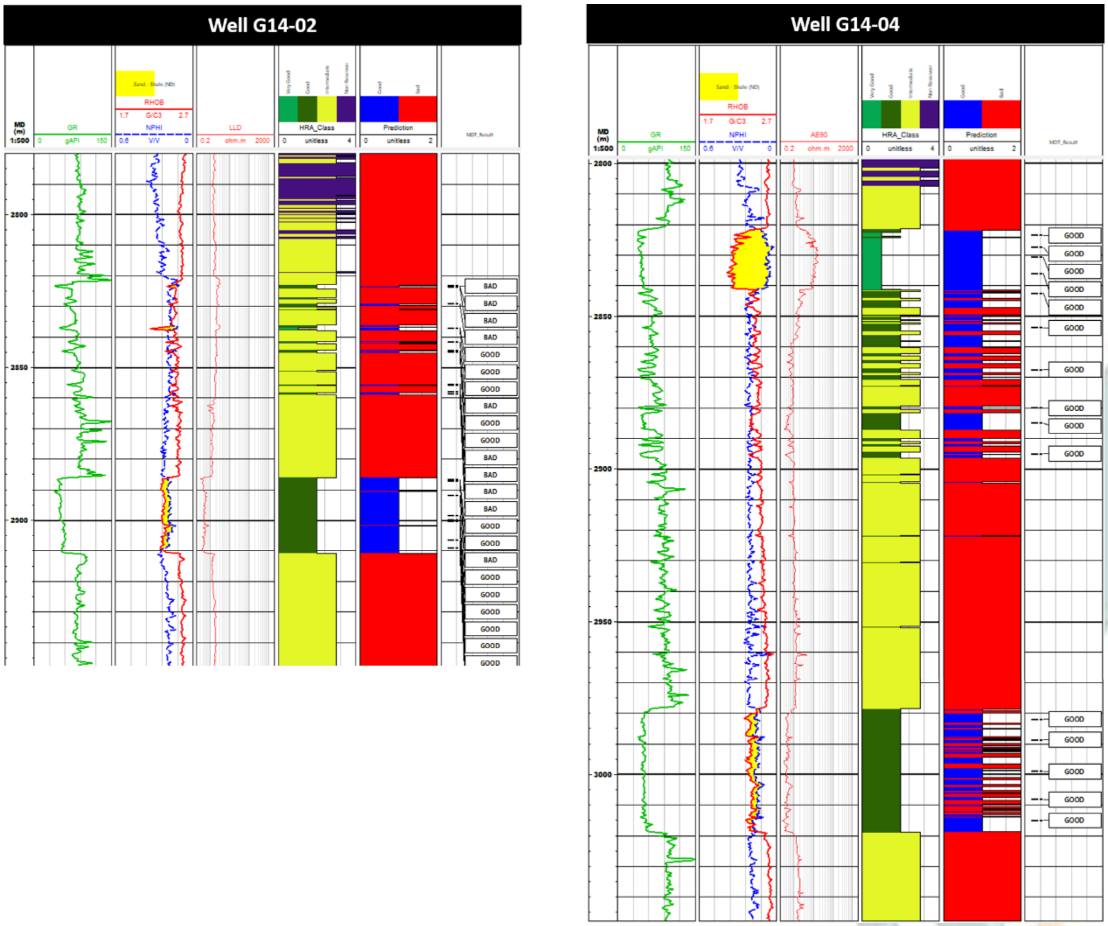


*Figure 2 Formation testing result using IPSOM for Well G14-02 and G14-04 as training wells*

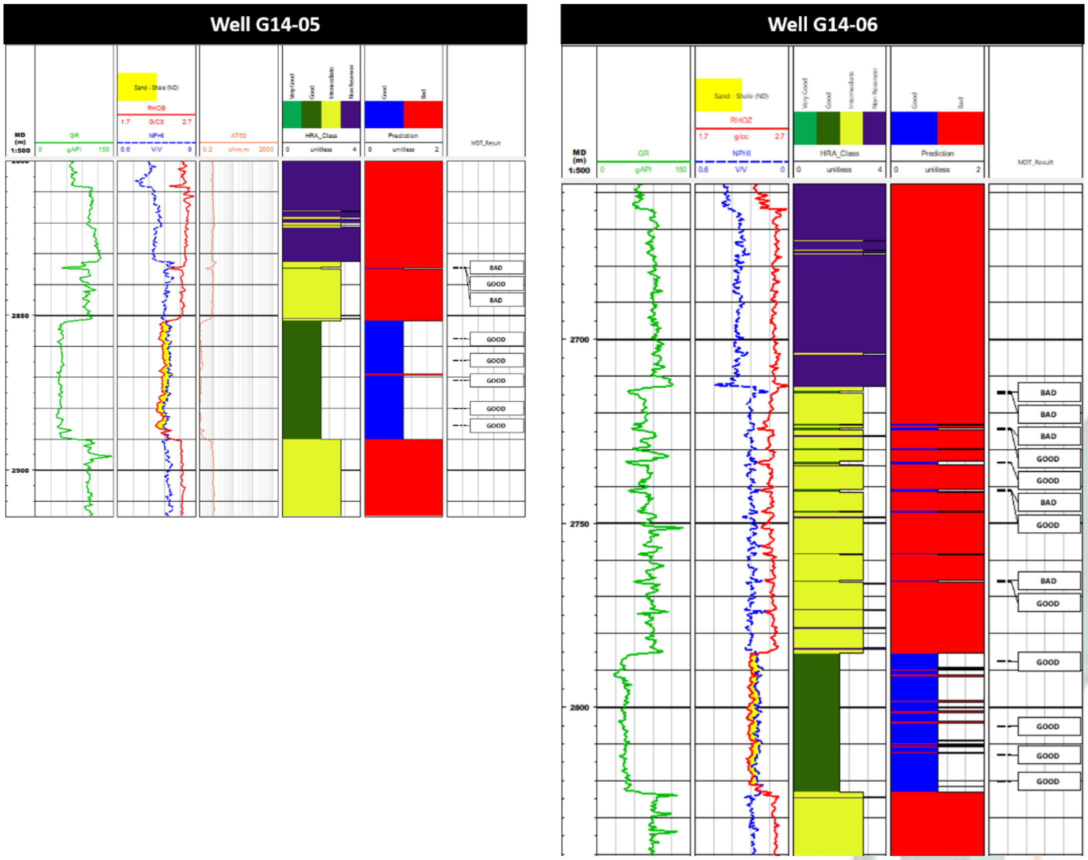


*Figure 3 Formation testing result using IPSOM for well G14-5 and G14-06 as training wells*

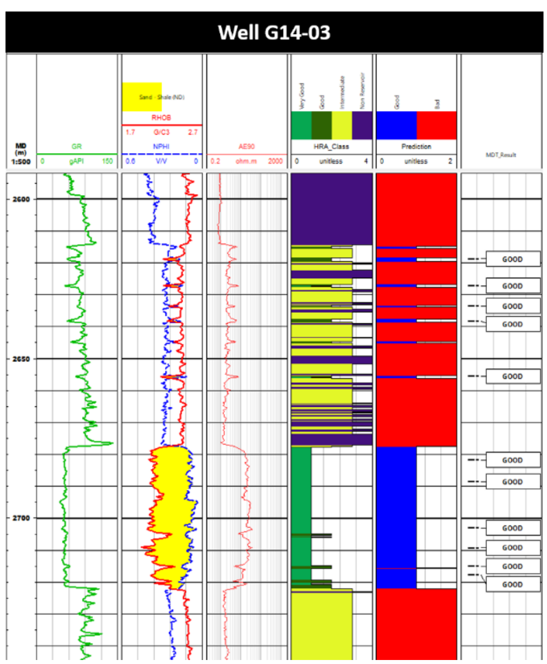


*Figure 4 Formation testing result using IPSOM for Well G14-03 as target well*

Several differences that can be observed of the WFT pretest status prediction between the ML-1 approach (ANN only) versus the ML-2 approach (IPSOM with ANN + SOM) are as follow. First, on the ML-2 approach, ranking of input variables based on correlation to the output are also computed. This helps to find the best set of variable input that can return best fitting model whereas the ML-1 approach would require manual iterations to include or exclude the irrelevant input variables. Second, the ML-2 approach automatically performed internal data pre-processing prior learning step and predict WFT pretest status for whole data interval, meanwhile, manual data pre-processing needs to be performed prior learning step for ML-1 approach.

## 4 Conclusions

In this study, there were two approaches that have been performed for predicting the WFT pretest status result using ML application. The first approach or ML-1 was scripted using python language with KMC and ANN method to classify the rock types and predict the WFT pretest status result. The second approach or ML-2 was performed using the built-in modules in Techlog™; Heterogeneous Rock Analysis (HRA) and Indexation Probability Self Organized Map (IPSOM) to classify the rock types and predict the WFT pretest status result. Both ML approaches were performed to classify the input data into four type of rock types with terminology of “Very Good”, “Good”, “Intermediate”, and “Non-Reservoir” rock types. Result would vary based on the samples of input data. On this case, both ML approaches were comparable and relatively consistent from one well to another. For the WFT pretest status result, both approaches show good accuracy (>80% accuracy) predicting the “train” and “target” wells dataset, with prediction accuracy of 86.7% and 100%, respectively, for ML-1 approach, and 83.2% and 100%, respectively for ML-2 approach.

In principle, there are many ways to perform machine learning to predict the WFT pretest status result. User can build from the scratch using python language with the suitable library (Scikit Learn, Keras, etc) and algorithm (K-Means Clustering, K-Nearest Neighbors, Support Vector Machine, Gaussian Naïve Bayes, etc) or user can build the workflow from the existing and build in machine learning module in Techlog ™ such as Heterogenous Rock Analysis (HRA), Indexation Probability Self Organized Map (IPSOM), Multiple Correspondences Analysis (MCA), K.mod and etc. Machine learning applications would require good data to learn from. It also requires good fundamental understanding of the domain itself, to state and achieve the objective of the workflow. A pre-defined model fitness criterion is necessary to drive the number of iterations required and parameter fine tuning in getting the best-fitting model for the case. Machine learning applications potentially help to extract the insight from vast offset wells database and get more consistent and reliable result in much faster manner than conventional or manual approach.

**References**

[1] Waldeland, A. U., Jensen, A. C., Gelius, L., and Solberg, A. H. S., Convolutional neural networks for automated seismic interpretation, The Leading Edge, 37, pp. 529-537.

[2] Hall, B., 2016. Facies classification using machine learning. The Leading Edge, 10, pp. 906-909.

[3] Doveton, J.D., 1994, Geologic Log Analysis Using Computer Methods; Second Edition, Published by the American Association of Petroleum Geologists (May 10, 1994).

[4] Handwerger, D., 2011, Scaling in Tight Gas Shale; Monograph of the First Shale Science Conference, Evolution of the Mental Picture of Tight Shales, Copernicus Science Center, Warsaw, Poland.

[5] Techlog ™ Help file, 'IPSOM', Schlumberger

[6] http://www.nlog.nl/en/boreholes