IATMI22-113

The Implementation of Supervised Learning and Cloud-Based Technology for Petrophysics: Identification of Hydrocarbon Prospect Zone and Classification of Rock Facies

Ilham Diaz Rahmat Nugroho*[1], Muhammad Destrayuda Trisna[2], and Muhammad Fajrul Haqqi[3]
[1,2]PERTAMINA HULU ROKAN, [3]UNIVERSITAS GADJAH MADA
* Email: diaz.ilhamrn@gmail.com

Abstract. The evolution of oil and gas industry technology has always been improving through time. The fluctuations, uncertainties, and investments had never been higher than the present. The combination of these conditions gave birth to the high risk and high return nature of Oil and Gas Industries, specifically in the Exploration and Production (E&P) activities. In E&P activities, one of the important analyses that has to be done was Petrophysics. Petrophysical analysis in E&P activities, were still using an archaic and lagging method with certain problems in prediction accuracy and rapid analysis. Thus, data-driven methods were expected to solve this challenge with machine learning (ML) techniques that could be applied to determine hydrocarbon prospect zones within the best acceptable accuracy range. In this study, the proposed solution utilizes data from Volve Oil Field, North Sea. This field potentially derive an advanced petrophysical model comparable to interpreted trends in seismic data and stratigraphic sequences due to their subsurface complexity. In this paper, variations of ML techniques have been implemented to build the models that predict the hydrocarbon prospect zone through petrophysical analysis and classify the rock facies. The proposed solution is examined through the evaluation process of 5 wells and compares the best algorithms of 4 options such as XGBoost, Extra Trees, and Random Forest, LightGBM. This solution contributes to reservoir characterization through the accomplished multi-well petrophysical analysis by high accuracy in just seconds. The prediction of petrophysical analysis obtained accuracy above 90% (including of permeability, water saturation, and effective porosity); rock facies reaches 80,0% (multiclass data up to 9 labels) using LightGBM Algorithm; meanwhile, the hydrocarbon prospect zone reaches 95,6% (binary data) using ExtraTreesClassifier algorithm. The processing time is calculated in a range of 0.2 - 1.5 seconds separated by training and testing data under an 8:2 ratio. Our models are in development using cloud computing technologies, accessible by nearly any device in any location, so long as there are internet connections available, giving the flexibility to utilize features previously unavailable online. The use of cloud computing and artificial intelligence technology, which is compatible and in sync with the development of this solution, can further increase the value, flexibility, and usability of primarily GGR data. The combination of the best efficiency and accuracy allows for better planning and execution of POD's, aid GGR engineers in finding prospect zones and locating recommended new well faster, which can further increase the economy, scalability, and accuracy of the hydrocarbon zone prediction.

Keyword(s): Petrophysics; Supervised Learning; Cloud Computing Technology; Rock Facies

1 Background (Mas Yuda)

1.1 Field Overview

Volve Oil Field is an Oil Field that is located in Norway. This oilfield is situated approximately 200 km to the east of Stavanger and 9 km from Sleipner Field. This field were operated by Equinor. The data from Volve field were released to public in June 2018.This field has an approximate water depth of 80 meter, and were considered as a shallow offshore field. Located in block 15/9 in the Central Part of North Sea, Volve Field has a complex structural configuration with several faults cutting the structure and causing the compartementalization of the reservoir. This field main target is the Hugin Formation which is estimated to contain around 78.6 million barrels of oil and 1.5 billion cubic meters of gas.Therefore, the complexity of the field and the completeness of the data available for downloads from Equinor gave birth to the decision of using this field as a good candidate for deployment of the algorithm.

1.2 Problem Identification

Upstream Oil and Gas Industry has a slow rate of technological improvement. Most of the technologies and the improvement revolves around how to acquire data faster and cheaper. But most people don't really invest in creating a more robust and more efficient way to process the acquired data. Moreover, the data itself (i.e. Well Logs) usually came as a bulk and the method to process the data had been proven to be slow and archaic. This problem came to reality as the nature of Oil and Gas Industry had been always known as High Risk and High Return. To answer this question, the deployment of Algorithms, Machine Learning has been always the most viable option for the researchers recently in order to create a more robust yet effective and efficient method to process upstream oil and gas data.

2 Methodology

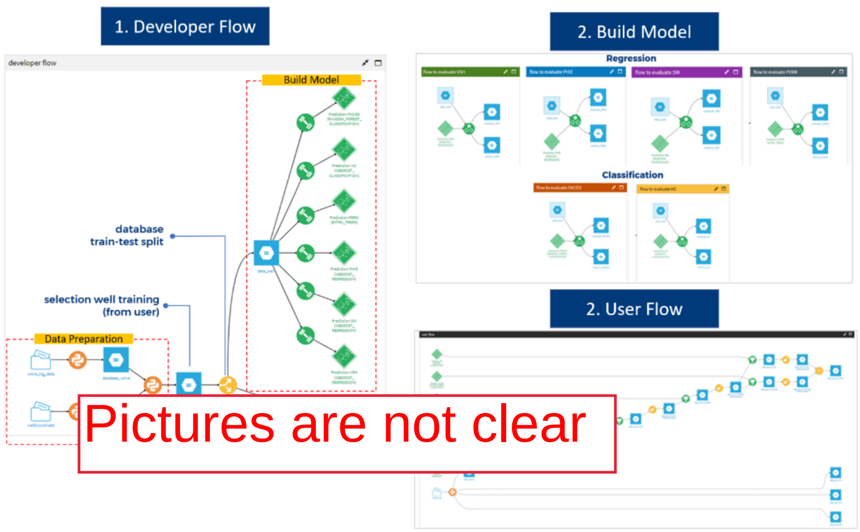

Figure 1. The flow of research steps.

The method in this research is divided into 3 stages: developer flow, which consists of data preparation, well training selection, and database creation. Then proceed to the build model, which consists of stages using regression and classification algorithms, and the last is user flow for the steps to use the developed web application. Fig. 1 shows the flow of research steps.

2.1 Build Machine Learning Model from Well Log Data

Creating the best model system with a fast data processing rate and the best accuracy value with minimum error. The user only enters input data, selects the pipeline as the output model, and gets the output data. The input data that the user owns can be in the form of petrophysical data or logging data, then well training is carried out on 5 wells and with 5 machine learning algorithms. The best prediction results will be used as the model output. Then this process also obtained output data in the form of petrophysical parameters, Lithofacies Classification, and Zone Prospect Identification. Fig. 2 shows a schematic for creating an ML model from well-log data

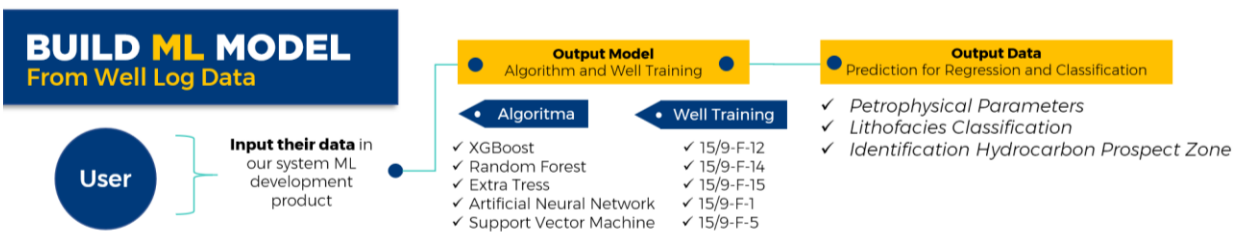

Figure 2. schematic for creating an ML Model from well-log data

2.2 Build The Spatial Data Analysis System

Build the best system for spatial data from well data location. Here our research will be recommended for the distance of the closest training well to the well data inputted. The existence of this system can assist in the accuracy of the ML model developed based on the distance between wells in the oil field. This process is divided into 2 outputs, namely schematic for creating an ML Model from well-log data and facies histogram to check the condition of the well training through histogram facies, well log, and petrophysical data analysis. The benefit of this system is that users can choose which well training will be correlated with the input well data. Fig. 3 shows the results of the well trajectory and facies histogram of the well log data.

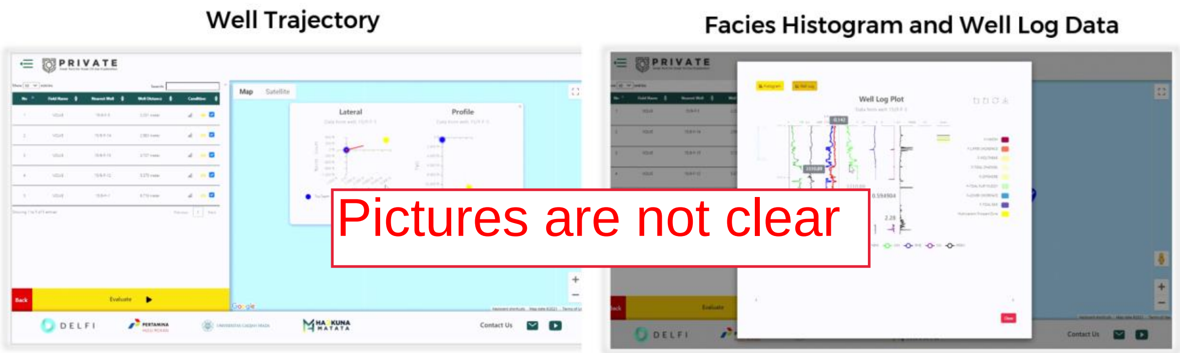

Figure 3. The well trajectory and facies histogram of the well log data.

2.3 Build The Database System

In this research, a database system was created. This system is very important because it will be used to process the system and store information systematically with good accuracy. This system will add benefits in the form of Grouping data for easy data identification, the database preparing the appropriate data with user requests for information quickly and right, Easy access, storage data, editing and deleting user data, and troubleshooting storage consuming conventional data very big space. Figure 4 shows the schema for building the database system in this research.

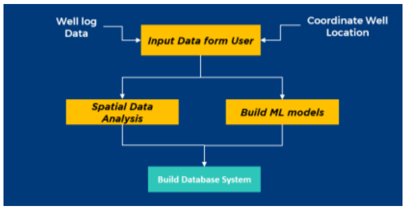

Figure 4. Schema building the database system

3 Result and Discussion

Each prediction generated is always given an evaluation through the error metrics of the algorithm used. Error metrics are used to determine how accurate the predictions generated from the algorithm used are to get the best model output. The smaller the error metrics value, the more accurate the predictions made. Then the evaluation is also reviewed from how big the scoring metrics are generated from the algorithm used, where the largest value is the most accurate value obtained. After that, it is also evaluated from the running time of the algorithm used, where the faster he processes the data, the more efficient the time needed to process the data. In this study, 5 algorithms were used to determine the petrophysical estimation. Fig. 5 shows each algorithm's comparison in evaluating the data processing results.

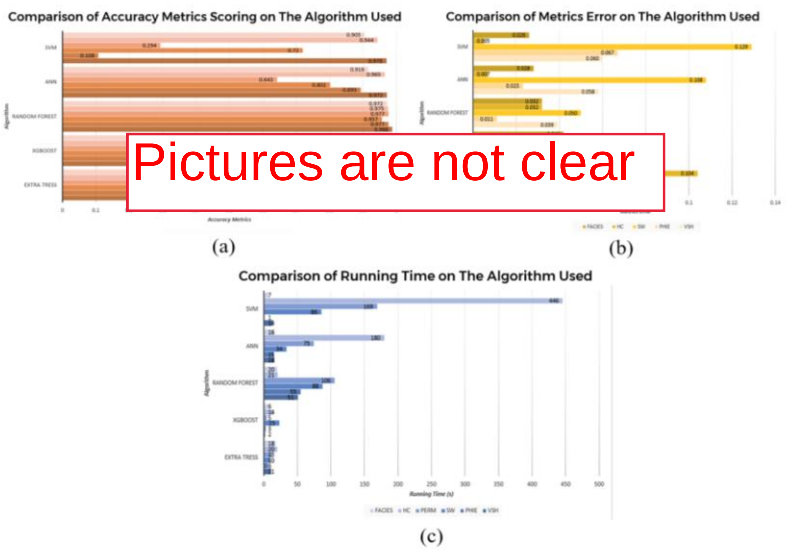

Figure 5. Comparison of evaluating machine learning model

In this research, a blind well test and Confusion Matrix were carried out to evaluate machine learning models so that they were accurate in getting predictions. Blind Well Test is a well log plot that represents the comparison between actual and predicted data in every mnemonic curve with evaluation metrics |error| on the regression model, and boolean (True or False) on classification model. To evaluate the machine learning model, we applied the model to a database with 35.902 data in 5 well with 20% test split. The best evaluation results are represented in the following graphs in Fig 6. Confusion Matrix is an N x N matrix used to evaluate a classification model's performance, where N is the number of target classes. The matrix compares the actual target values with those predicted by the machine learning model.

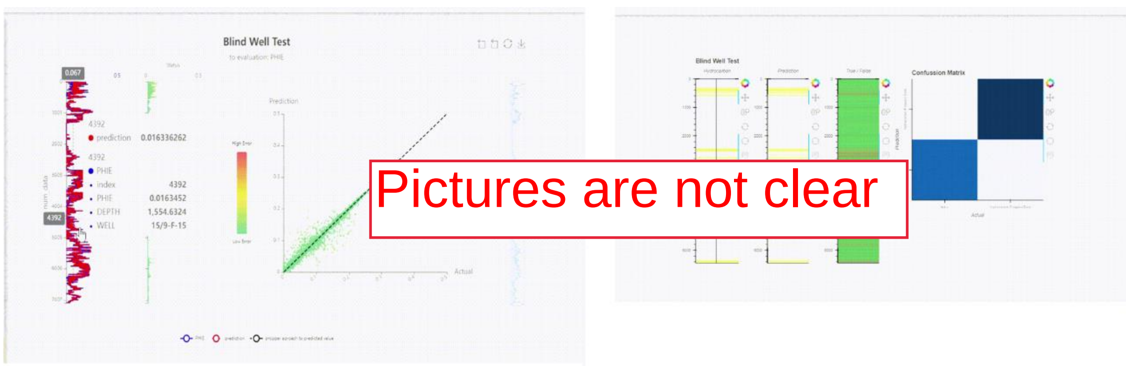

Figure 6. Blind well test and Confusion Matrix Result

4 Conclusion

1. Prediction of petrophysical analysis obtained accuracy above 90% (covering permeability, water saturation, and effective porosity); rock facies reached 80.0% (multiclass data up to 9 labels) using the LightGBM algorithm; while the hydrocarbon prospect zone reached 95.6% (binary data) using the ExtraTreesClassifier algorithm.
2. Efficient data processing time ranges from 0.2-1.5 seconds separated by training and test data under a ratio of 8:2.
3. The combination of the best efficiency and accuracy allows for better planning and execution of POD's aid GGR engineers in finding prospect zones, and locating recommended new well faster, which can further increase the economy, scalability, and accuracy of the hydrocarbon zone prediction.

References

[1] Dwihusna N 2020 Seismic And Well Log Based Machine Learning Facies Classification In The Panoma-Hugoton Field, Kansas And Raudhatain Field, North Kuwait. *Thesis. Department of Geophysics*, (Colorado: Colorado School of Mines).
[2] Chen T and Guestrin C 2016 XGBoost: A scalable tree boosting system. *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp 785–794.
[3] Friedman J H 2001 Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp 1189–1232.
[4] Hall B 2016 Facies classification using machine learning. *The Leading Edge*, chapter 35 vol 10 pp 906–909.
[5] Zheng Z, Pan S, Luo H and Guo Z 2020 Porosity prediction based on GS+ GA-XGBoost.