Scientific Contributions Oil & Gas, Vol. 47. No. 1, April: 21 - 30

SCIENTIFIC CONTRIBUTIONS OIL AND GAS

Testing Center for Oil and Gas

LEMIGAS

Journal Homepage:http://www.journal.lemigas.esdm.go.id

ISSN: 2089-3361, e-ISSN: 2541-0520

Comparison of Facies Estimation of Well Log Data Using Machine Learning

Arya Dwi Candra[1], Pradini Rahalintar[1], Sulistiyono[1] and Urip Nurwijayanto Prabowo[2]

[1]Oil and Gas Production Engineering Department, Politeknik Energi dan Mineral Akamigas
Gadjah Mada Street No. 38, Cepu, 58315, Indonesia

[2]Department of Physics, Universitas Jenderal Soedirman
dr. Soeparno street No. 61, Purwokerto, 53123, Indonesia

Corresponding author: arya.candra@esdm.go.id

Manuscript received: August 18st, 2023; Revised: September 26st, 2023
Approved: January 25st, 2024; Available online: February 20st, 2024

ABSTRACT - Accurately identifying lithological facies is crucial for comprehending geological variations in a proven reservoir. To enhance the accuracy of facies classification compared to previous studies on the same dataset, five distinct machine learning algorithms were employed to predict facies in both a panoma field dataset and Z-Field, Indonesia. The analysis data samples with known facies, originating from core data from Panoma Field and Z-Field. Facies classification was addressed using five well-known classification algorithms, namely K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Neural Network Classifier (NNC), Random Forest Classifier (RFC), and Decision Tree Classifier (DTC). The dataset was divided into training and testing subsets to evaluate the machine learning models. The five suggested algorithms demonstrate effective facies prediction, closely aligning with the actual facies in the test wells within the Panoma field. However, these algorithms struggle to predict facies accurately in the Z field well, primarily attributed to the imbalanced data distribution between sandstone-claystone and siltstone-limestone. Equalizing the number of facies labels in the training data becomes essential to enable the algorithm to recognize patterns and accurately estimate all facies types.

Keywords: facies, well logging, machine learning, supervised learning.

How to cite this article:

Arya Dwi Candra, Pradini Rahalintar, Sulistiyono and Urip Nurwijayanto Prabowo, 2024, Comparison of Facies Estimation of Well Log Data Using Machine Learning, Scientific Contributions Oil and Gas, 47 (1) pp. 21-30. DOI.org/10.29017/SCOG.47.1.1593.

INTRODUCTION

In general, geologists often rely on direct core measurements to discern layer boundaries, enabling the acquisition of detailed well information. Nevertheless, this approach proves inefficient, costly, and time-consuming when it comes to analysis and interpretation. Well-log analysis emerges as a highly valuable method for characterizing well lithology, encompassing parameters like pressure, saturation, fluid type, pore characteristics, and geometry. Traditional techniques employed in well-log data interpretation, such as visual inspection, primarily hinge on the interpreter's experience, yielding varying opinions and interpretations. To circumvent the issues arising from divergent interpretations, as well as to enhance cost-effectiveness and efficiency, numerous computational algorithms have been embraced as a viable and cost-effective alternative.

Facies classification is the process of identifying rock lithology based on indirect measurements such as well log measurements. Facies classification is a process that is generally carried out manually by experienced geologists so it takes a relatively long time and is inefficient (Saroji et al., 2021). Based on this, many machine learning applications have been carried out on well log data to perform facies classification automatically (Al-Mudhafar, 2017; Al-Mudhafar et al., 2022; Hall, 2016; Hardanto & Wulandhari, 2021; Imamverdiyev & Sukhostat, 2019; Mandal & Rezaee, 2019; Merembayev et al., 2021; Mohamed et al., 2019; Pratama et al., 2020; Saroji et al., 2021; Singh et al., 2020)an integrated procedure was adopted to obtain accurate lithofacies classification to be incorporated with well log interpretations for a precise core permeability modeling. Probabilistic neural networks (PNNs. Machine Learning (ML) is a scientific discipline that defines statistical or mathematical models based on data (Dramsch, 2019). Machine learning applications can increase the effectiveness and efficiency in finding solutions to geophysical interpretation problems on complex data (Dell'Aversana, 2019; Merembayev et al., 2021) and reservoir quality assessment with lower cost and time (Dixit et al., 2020).

In this study, facies classification was carried out in the Panoma Field dataset and Z field, Indonesia using the support vector machine (SVM), K-Nearest Neighbor (KNN), Neural Network Classifier (NNC), Random Forest Classifier (RFC), and Decision Tree Classifier (DTC).

METHODOLOGY

Machine Learning Algorithm

At the cutting edge of artificial intelligence (AI) technology, machine learning comprises a set of data analysis algorithms covering classification, regression, and clustering methods (Hall, 2016). This technology is broadly divided into supervised and unsupervised categories. In supervised machine learning, essential elements involve input features and target output. In the field of geoscience, machine-led applications utilizing wire-line logs are commonly employed to tackle challenges in oil and gas exploration, development, and production environments.

This present research focuses on improving the recognition of lithological facies in a published dataset by employing supervised classification algorithms. The objective of this inquiry is to assess the potential of these algorithms in enhancing the identification of lithological facies within the realm of geoscience.

The research consists of 4 stages, namely data selection, training phase, verification, and validation. The data selection phase began by selecting wells as training data and test data. The training data from the wells is then separated as much as 80% for the training phase and 20% for the verification stage. The data parameters used in this study can be seen in Figure 1.

In the training phase, the machine learning algorithm will recognise and form a relationship pattern between input and output. In this study, the input used is 80% training data from wireline logging measurement results and the output used is facies data.

The verification stage is the stage of knowing the accuracy of the machine learning training results. This stage uses 20% of the training data that is not used in the training phase. The machine learning algorithm makes predictions on this data and the results are matched with existing facies data based on the confusion matrix. The parameter data used in this study can be seen in Figure 1. The training and test data were normalized so that the input features have a uniform scale to help in faster convergence of the algorithm. The validation phase is a stage to validate the ability of machine learning to estimate using different data sets. If the previous stage uses data from the same well, then the validation stage uses data from different wells, namely test data. The results of the facies prediction will be compared with the existing facies data to determine the accuracy of the estimation. In this investigation, the algorithm's accuracy is assessed using the Jaccard Index and F1-Scores. The Jaccard Index, also known as the Jaccard similarity coefficient, serves as a statistical measure to gauge the similarities between sets of samples. This metric highlights the similarity between finite sample sets and is formally expressed as the size of the intersection divided by the size of the union of the two labeled sets, as outlined in Equation 1.

$$J(y, y1) = \left|\frac{y \cap y1}{y \cup y1}\right| \quad (1)$$

The F1 score is computed as the harmonic mean of precision and recall. An F1 score attains its optimal value of 1 when precision and recall are perfect, and it diminishes to 0 at its worst. Equation 2 illustrates the calculation of accuracy using F1 scores.

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (2)$$

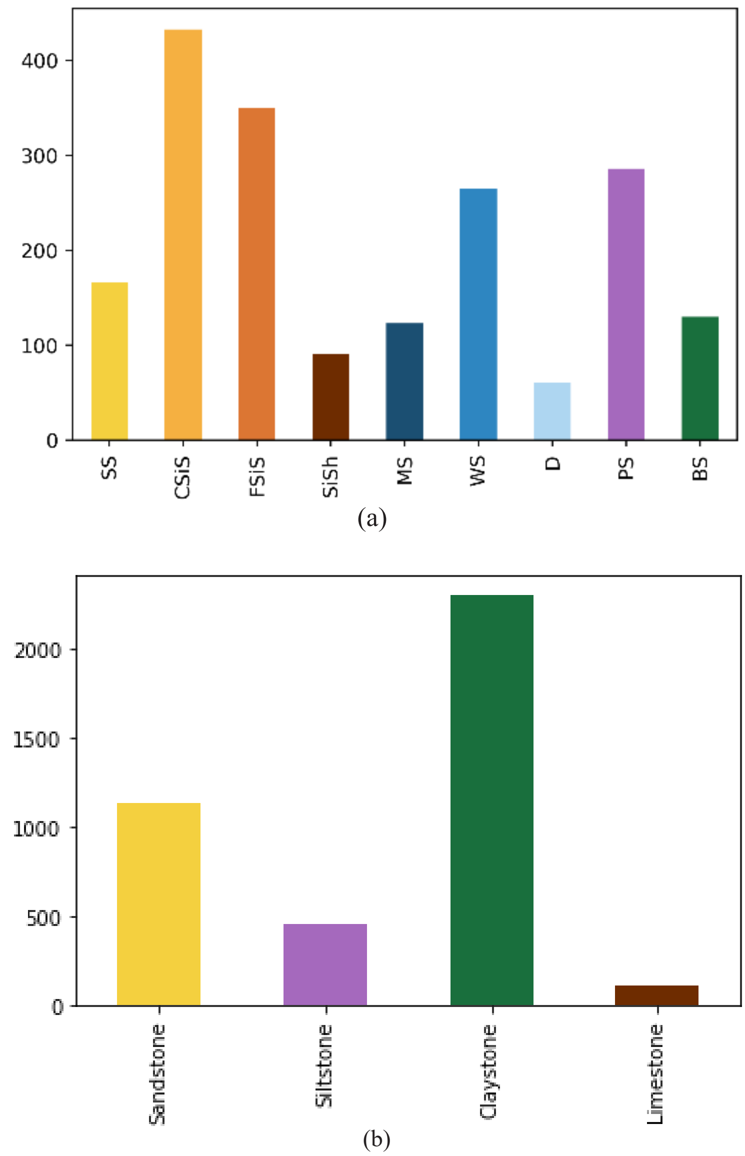
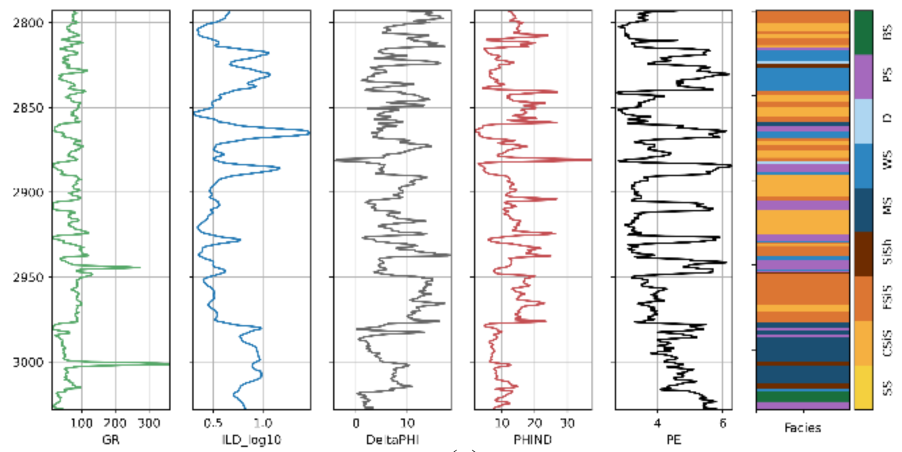


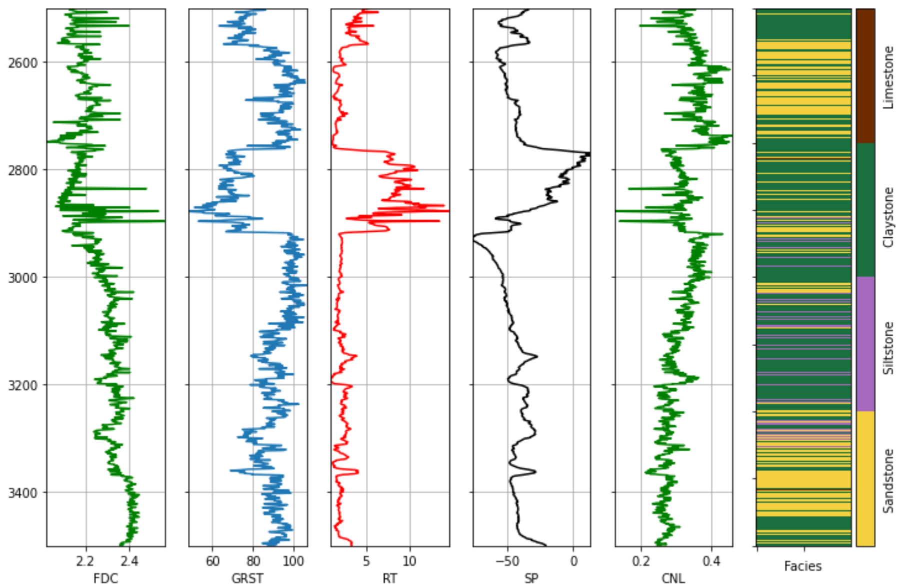
Figure 1
Distribution of facies on training data. (a) Panoma field and (b) Z Field.

Data Preparation

The data for this study using 2 field, Panoma (Dubois et al., 2007)seven classifiers based on four different approaches were tested in a rock facies classification problem: classical parametric methods using Bayes' rule, and non-parametric methods using fuzzy logic, k-nearest neighbor, and feed forward-back propagating artificial neural network. Determining the most effective classifier for geologic facies prediction in wells without cores in the Panoma gas field, in Southwest Kansas, was the objective. Study data include 3600 samples with known rock facies class (from core and Z Field (Prabowo et al., 2023). The input dataset for Panoma field is constituted by a suite of five wireline log curves collected across the nine wells, encompassing average neutron density porosity (PHIND), neutron density porosity difference (DeltaPHI), , photoelectric effect (PE), resistivity (ILD_log10) and gamma-ray (GR). The lithological composition of the target stratum is characterized by nine distinct classes, meticulously delineated by (Dubois et al., 2007)seven classifiers based on four different approaches were tested in a rock facies classification problem: classical parametric methods using Bayes' rule, and non-parametric methods using fuzzy logic, k-nearest neighbor, and feed forward-back propagating artificial neural network. Determining the most effective classifier for geologic facies prediction in wells without cores in the Panoma gas field, in Southwest Kansas, was the objective. Study data include 3600 samples with known rock facies class (from core through an exhaustive petrophysical assessment and core analysis. The nomenclature for facies classification adheres to established categories: Phylloid-algal bafflestone (BS) , Packstone-grainstone (PS), Dolomite (D), Wackstone (WS), Mudstone (MS), Marine siltstone and shale (SiSh), Non-marine fine siltstone (FSiS), Non-marine coarse siltstone (CSiS), and Non-marine sandstone (SS). The input dataset for Z field is constituted by a suite of five wireline log curves collected across the three wells, encompassing density (FDC), gamma ray (GRST), resistivity (RT), SP (SP) and Compensated Neutron Log (CNL). The facies labels used in the well data include (1) Sandstone, (2) Siltstone, (3) Claystone and (4) Limestone. Figure 2 delineates the facies log of a training well in conjunction with GR logs. The dataset is enriched with two crucial geological constraining variables: nonmarine and marine identifiers (NM and M) and relative position with respect to the base of formation (NM and M), denoted as RELPOS. Figure 3 presents a 1-D histogram of petrophysical logs, colour-coded with each facies in the diagonal orientation, while off-diagonal elements represent cross-plots of input features (well-logs). In Figure 3, every panel displays the correlation between two variables plotted on the x and y axes. A stacked bar plot on the diagonal exhibits the distribution of each data point, and the colour of each point corresponds to its respective facies.



(a)



(b)

Figure 2
Wireline log and Facies distribution of Panoma field. (a) Panoma field and (b) Z Field.

RESULT AND DISCUSSION

Panoma Field

Analyzing data, comprehending the statistical representation of data samples, and visualizing the correlation function between input and output are pivotal aspects of any successful machine learning endeavor. The process of data preparation and outlier detection commenced with a visual examination of 1-D diagrams depicting input features, such as histograms where each facies is distinguished by color coding, as illustrated in Figure 3. Notably, the off-diagonal plots in a 2-dimensional space (Figure 3) demonstrate that no singular feature can segregate facies linearly.

Porosity and density logs stand out as physically significant properties for reservoir characterization, playing a crucial role in rock sample identification. To enhance the dataset, DeltaPHI and PHIND from the input dataset undergo simple arithmetic operations, resulting in the creation of two distinct porosity logs (PHID and NPHI).

Given the requirement for consistent input feature scaling in machine learning applications, each feature undergoes transformation through a mean normalization operator. This procedural step facilitates a faster convergence of the algorithm. To optimize various hyperparameters of the selected machine learning algorithm—such as the number of iterations, regularization value, gamma, and others—a validation dataset is employed. This ensures a meticulous tuning of parameters for improved model performance.

The remaining sample points are split into training and validation sets at an 80:20 ratio, respectively. Five supervised learning algorithms (SVM, KNN, NNC, RFC, and DTC) are utilized on the training dataset to acquire the mapping function between input features and the output facies class. The facies under consideration are not strictly distinct; certain instances display a blending, posing a challenge to achieve complete accuracy in classification.. The accuracy of specific facies is assessed on a cross-validation dataset to determine the most effective classifier model. SVM, KNN, NNC, RFC, and DTC algorithms predict classification accuracy using the Jaccard Index and F1 Scores, as presented in Table 1. The Jaccard Index indicates predicted classification accuracies of 55%, 52%, 54%, 50%, and 57% for SVM, KNN, NNC, RFC, and DTC, respectively. In terms of F1 Scores, the predicted classification accuracies are 48%, 53%, 49%, 53%, and 47%, respectively. These results confirm that the mentioned classifiers can be chosen for facies log determination in the Panoma field. Figure 4 illustrates the close alignment between the predicted facies and the actual facies class in the test well.

Table 1
Validation data of test data in panoma field

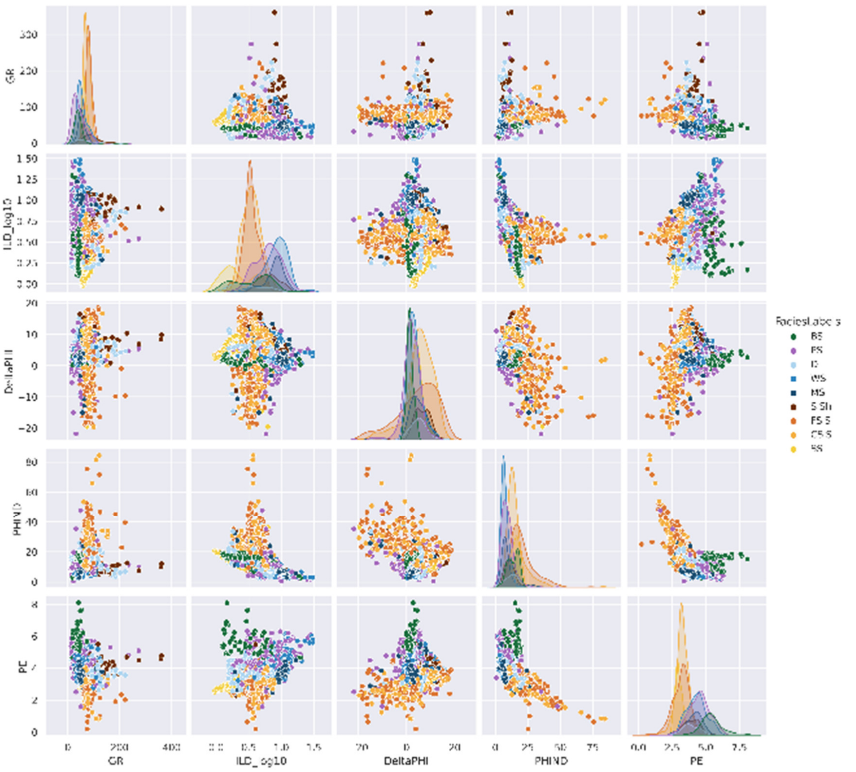
Machine Learning Model	Jaccard Index	F1 Score
SVM	0.55	0.48
RFC	0.50	0.53
NNC	0.54	0.49
KNN	0.52	0.53
DT	0.57	0.47

Z Field

This study uses log data from the Z field, Indonesia. This data has 3 wells data parameters (input) from wireline logging measurements and 1 facies label (output). The facies data is the result of geological interpretation based on rock samples during drilling and petrographic analysis. Well data from wireline logging measurements used include density (FDC), gamma ray (GRST), resistivity (RT), SP (SP) and Compensated Neutron Log (CNL). The facies labels used in the well data include (1) Sandstone, (2) Siltstone, (3) Claystone and (4) Limestone.

The accuracy of five algorithm facies estimation on the test data (validation stage) is determined based on the Jaccard Index and F1 Score as seen in Table 2. Based on the result, the predicted classification using the Jaccard Index for SVM, KNN, NNC, RFC, and DTC algorithm are 48%, 52%, 37%, 28% and 38%, respectively. While using F1 Scores, the predicted classification accuracy are 56%, 52%, 66%, 71% and 64%, respectively. Figure 5 show the matched among the actual facies and predicted facies on the test well.

Figure 5 also shows that the algorithm has not been able to estimate the siltstone and limestone facies. This is likely due to the small amount of siltstone and limestone facies training data (Figure 1b) that makes the algorithm unable to recognize the pattern of the facies output. Training data should be data with an even distribution of the number of facies labels.



(a)

(b)

Figure 3
A pair-plot graph depicting histogram distributions of petrophysical logs along the diagonal, while the off-diagonal section illustrates two-dimensional cross-plots for every log combination. Each facies is distinguished by a unique color to enhance clarity in visualization. The fields presented include (a) Panoma Field and (b) Z Field.

| DOI.org/10.29017/SCOG.47.1.1593

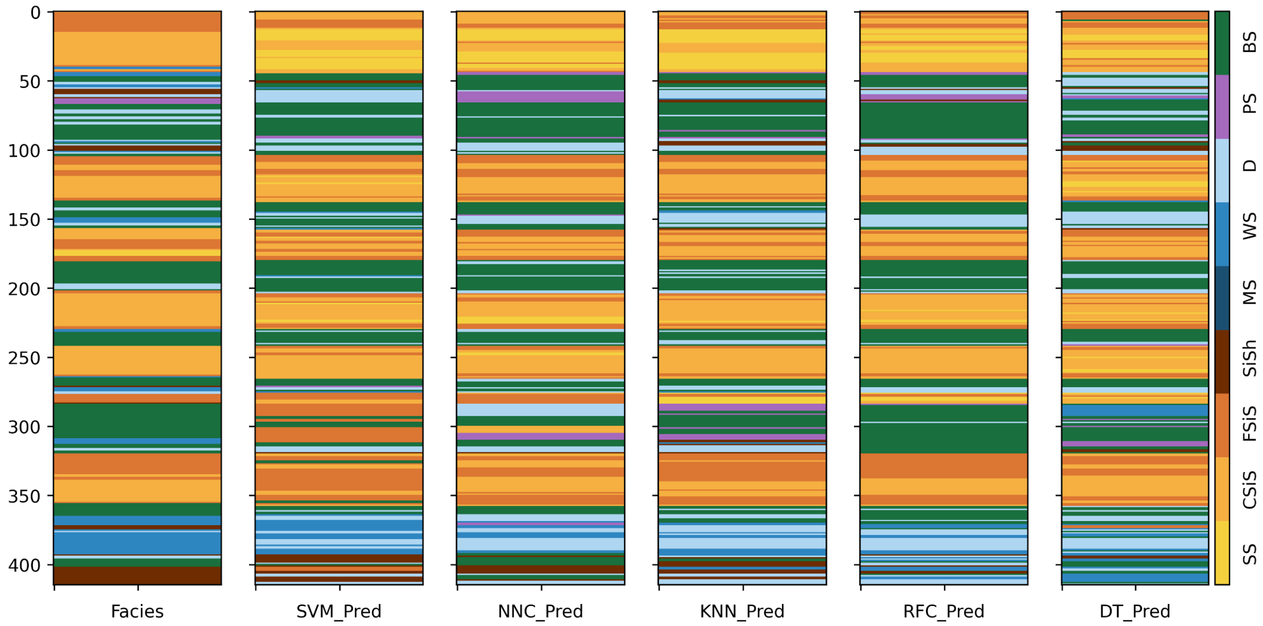


Figure 4
Various model predictions using machine learning

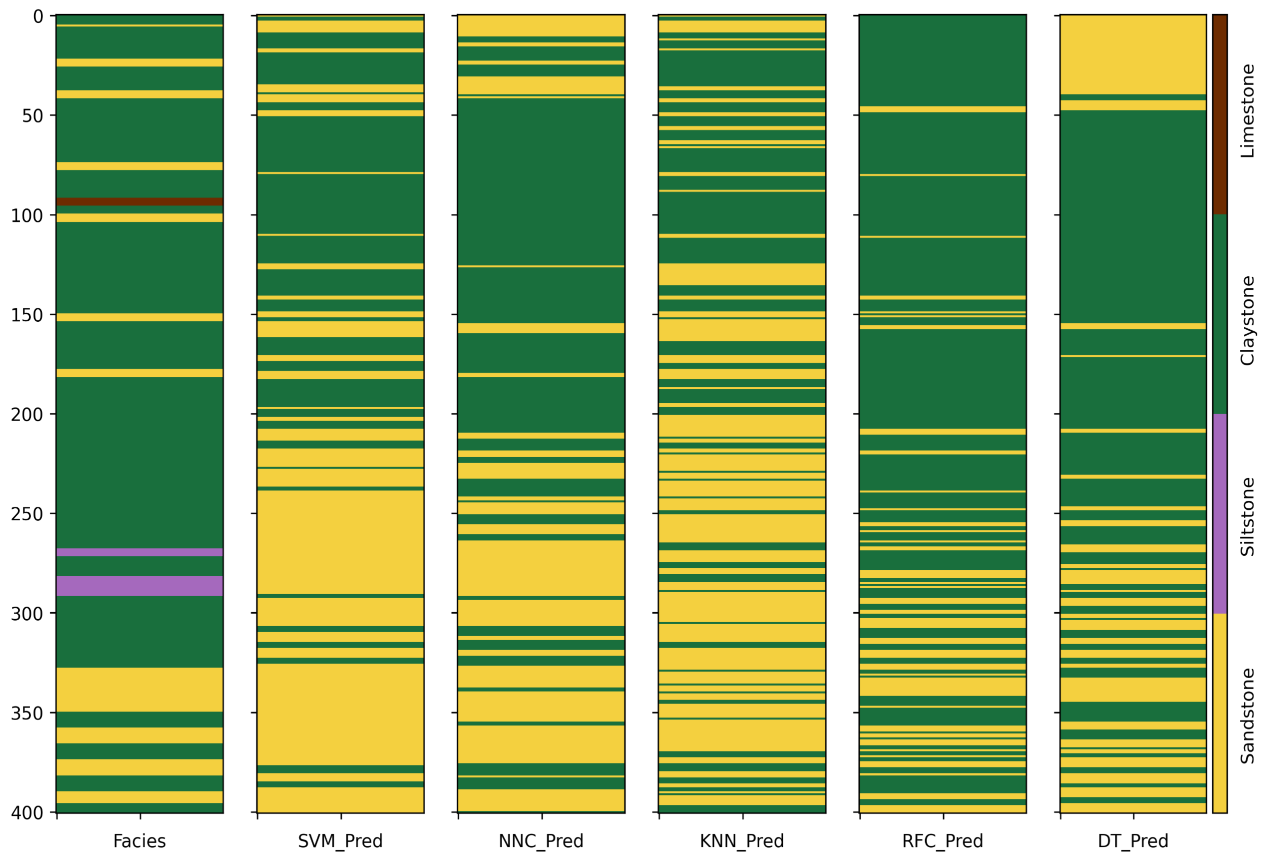


Figure 5
Various model predictions using machine learning

Comparison of the estimated facies results and the actual facies can be seen in Figure 5. It can be seen that the five algorithms fail to recognize the input-output pattern in the training data so that it dominantly estimates the facies as claystone and Sandstone. Claystone and Sandstone have the most distribution of facies in the training data (Figure 1b).

At the training stage, machine learning parameters must be determined correctly because they greatly affect the results of prediction accuracy. The results of facies estimation using the algorithm have better accuracy than SVM. However, the two algorithms have not been able to determine the Siltstone and Limestone facies due to the small number of these facies in the training data so the algorithm has not recognized the facies input-output pattern yet. An even distribution of facies labels on the training data is needed so that the algorithm can recognize patterns and can estimate all types of facies correctly

Table 2
Validation data of test data in Z field

Machine Learning Model	Jaccard Index	F1 Score
SVM	0.48	0.56
RFC	0.28	0.71
NNC	0.37	0.66
KNN	0.52	0.52
DT	0.38	0.64

The accuracy of machine learning-based facies prediction is heavily influenced by factors such as the quality and volume of input data, parameters in the training dataset, and the specific machine learning algorithm employed (Dixit et al., 2020). An immediate observation reveals that certain facies are disproportionately represented, leading to a significant imbalance in the classification problem. Notably, Siltstone and limestone have fewer samples in the dataset, explaining why the classifier struggled to effectively characterize them. Conversely, Sandstone and Claystone, benefiting from a greater number of samples, were among the facies most accurately detected.

CONCLUSIONS

The facies estimation results using the five algorithms have different accuracies between the Panoma and Z fields. The five proposed algorithms can predict facies with results close to the actual facies of the test wells in the Panoma field. Whereas the five proposed algorithms are not able to predict facies with results close to the actual facies from the Z field well. This is due to the unequal amount of data between sandstone-claystone and siltstone-limestone in the Z field. It is necessary to equalize the number of facies labels in the training data so that the algorithm can recognize patterns and estimate all facies types correctly.

ACKNOWLEDGEMENTS

We would like to thank Politeknik Energi dan Mineral Akamigas for their support and permit to use facilities of the Oil and Gas Upstream Laboratorium.

GLOSSARY OF TERM

Symbol	Definition
AI	Artificial Inteligent
BS	Phylloid-algal bafflestone
CNL	Compensated Neotron Log
CSiS	Non-marine coarse siltstone
D	Dolomite
DelthaPHI	Neutron density porosity difference
DTC	Decision Tree Classifier
F1-Score	Measure of the harmonic mean of precision and recall
FDC	Encompassing Density
FSiS	Non-marine fine siltstone
GR	Gamma Ray
ILD_log10	Resistivity Log
KNN	K-Nearest Neighbour
M	Marine
MS	Mudstone
NM	Non Marine

NNC	Neural Network Classifier
PE	Photoelectric Effect

REFERENCES

Al-Mudhafar, W. J. (2017). Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms. *Journal of Petroleum Exploration and Production Technology*, *7*(4), 1023–1033. https://doi.org/10.1007/s13202-017-0360-0

Al-Mudhafar, W. J., Abbas, M. A., & Wood, D. A. (2022). Performance evaluation of boosting machine learning algorithms for lithofacies classification in heterogeneous carbonate reservoirs. *Marine and Petroleum Geology*, *145*, 105886. https://doi.org/10.1016/j.marpetgeo.2022.105886

Dell'Aversana, P. (2019). Comparison of different Machine Learning algorithms for lithofacies classification from well logs. *Bollettino Di Geofisica Teorica Ed Applicata*, *60*(1).

Dixit, N., McColgan, P., & Kusler, K. (2020). Machine Learning-Based Probabilistic Lithofacies Prediction from Conventional Well Logs: A Case from the Umiat Oil Field of Alaska. *Energies*, *13*(18), Article 18. https://doi.org/10.3390/en13184862

Dramsch, J. S. (2019). *Machine learning in 4D seismic data analysis: Deep neural networks in geophysics*.

Dubois, M. K., Bohling, G. C., & Chakrabarti, S. (2007). Comparison of four approaches to a rock facies classification problem. *Computers & Geosciences*, *33*(5), 599–617. https://doi.org/10.1016/j.cageo.2006.08.011

Hall, M. (2016). A user guide to the geophysical tutorials. *The Leading Edge*, *35*(2), 190–191. https://doi.org/10.1190/tle35020190.1

Hardanto, L. T., & Wulandhari, L. A. (2021). Lithofacies Classification Using Supervised and Semi-Supervised Machine Learning Approach. *International Journal on Advanced Science, Engineering and Information Technology*, *11*(2), 542. https://doi.org/10.18517/ijaseit.11.2.11764

Imamverdiyev, Y., & Sukhostat, L. (2019). Lithological facies classification using deep convolutional neural network. *Journal of Petroleum Science and Engineering*, *174*, 216–228. https://doi.org/10.1016/j.petrol.2018.11.023

Mandal, P. P., & Rezaee, R. (2019). Facies classification with different machine learning algorithm – An efficient artificial intelligence technique for improved classification. *ASEG Extended Abstracts*, *2019*(1), 1–6. https://doi.org/10.1080/22020586.2019.12072918

Merembayev, T., Kurmangaliyev, D., Bekbauov, B., & Amanbek, Y. (2021). A Comparison of Machine Learning Algorithms in Predicting Lithofacies: Case Studies from Norway and Kazakhstan. *Energies*, *14*(7), Article 7. https://doi.org/10.3390/en14071896

Mohamed, I. M., Mohamed, S., Mazher, I., & Chester, P. (2019). Formation Lithology Classification: Insights into Machine Learning Methods. *Day 2 Tue, October 01, 2019*, D021S033R005. https://doi.org/10.2118/196096-MS

Prabowo, U. N., Ferdiyan, A., Raharjo, S. A., Sehah, S., & Candra, A. D. (2023). Comparison Of Facies Estimation Using Support Vector Machine (SVM) And K-Nearest Neighbor (KNN) Algorithm Based On Well Log Data. *Aceh International Journal of Science and Technology*, *12*(2), Article 2. https://doi.org/10.13170/aijst.12.2.28428

Pratama, H., Syahputra, L. A., Albany, M. F., Abdullah, A., Suhardja, S. K., Kusumah, E., Astuti, W., & Mujihardi, B. (2020). Early Results of Comparison between K-Nearest Neighbor and Artificial Neural Network Method for Facies Estimation. *Jurnal Geofisika*, *18*(1), 7–13.

Saroji, S., Winata, E., Pratama, P., Hidayat, W., Prakoso, S., & Herdiansyah, F. (2021). The implementation of machine learning in lithofacies classification using multi well logs data. *Aceh Int. J. Sci. Technol*, *10*, 9–17.

Singh, H., Seol, Y., & Myshakin, E. M. (2020). Automated Well-Log Processing and Lithology Classification by Identifying Optimal Features

Through Unsupervised and Supervised Machine-Learning Algorithms. *SPE Journal*, *25*(05), 2778–2800. https://doi.org/10.2118/202477-PA.