International Journal on Advanced Science Engineering Information Technology

Vol.11 (2021) No. 2
ISSN: 2088-5334

Lithofacies Classification Using Supervised and Semi-Supervised Machine Learning Approach

Lilik T. Hardanto[a,*], Lili Ayu Wulandhari[a]

[a]*Department of Computer Science, Binus University, Jakarta, 11530, Indonesia*
*Corresponding author: *lilik.hardanto@binus.ac.id*

***Abstract*— The machine learning approach can help Geoscientists do their work in well log analysis to developing the oil and gas field. Prediction categorical or numerical response variable using a set of predictor variables supervises and semi-supervises learning is an important goal of the machine learning approach in classifying lithofacies using well log data. Semi-supervised classification offers the possibility of exploring the structure of the data without entirely external knowledge or guidance in the form of target or class information, and semi-supervised is very rarely research in the field of lithofacies classification. Well log data in gamma-ray, resistivity, neutrality, and density logs are collected and selected for data processing and transformation. The use of machine learning algorithms such as Naïve Bayes, SVM, and Decision Tree is to find the log pattern or pattern classifications of lithofacies in supervised and semi-supervised to create a model with conditions requiring the change of data and the corresponding requirements. All supervised machine learning algorithms have the best accuracy because algorithms provide useful predictive in classifications based on the target but not if there are no targets given or semi-supervised. This paper compares some of the famous classification algorithms of machine learning, such as Decision tree, SVM, and Naïve Bayes, on classifying lithofacies with supervised and semi-supervised learning. This research found that the semi-supervised learning of Naïve Bayes has performed well in classified lithofacies. In contrast, in supervised learning, Decision Tree and SVM are superior in accuracy and visualization approach based on expert's interpretation.**

***Keywords*— Machine learning; decision tree; SVM; naïve bayes; lithofacies; supervised; semi-supervise learning.**

Manuscript received 28 Jun. 2020; revised 6 Jul. 2021; accepted 2 Feb. 2021. Date of publication 30 Apr. 2021.

I. INTRODUCTION

Generally, machine learnings used to predict the probability of hazards such as in engineering, earthquakes, and weather forecasts also predict the dangers. Many researchers have been conducting to predict and classify lithofacies using well log curves data [1 that use CNN (Convolutional Neural Network) [2], [3], Random Forest [4], [5], Neural Network, and Ad boost [6], KNN clustering [7], Gradient boosting classifier [8] algorithms.

Predicting a categorical or numeric response variable using a full predictor variable set or supervised and partially or semi-supervised learning is the primary goal of machine learning. Supervised classification offers the possibility of exploring the data structure with any external knowledge or guidance in target or class information and often reveals features that are expected using the expert bias. In general, supervised learning in machine learning is widely used and researched to conduct data classification based on the specified target or label. The predictive model is given a bright instructional start from the beginning, like learning and how the historical data learned.

Semi-supervised classification offers the possibility of exploring the structure of the data without entirely external knowledge or guidance in the form of target or class information. It often reveals features that were and not expected, avoiding the total expert bias which supervised methods build. Semi-supervised is very rarely research in the field of lithofacies classification.

Many researchers have investigated the technique of combining the predictions of multiple classifiers to produce a single classifier. Ensemble machine learning consists of a combination of multiple Artificial Intelligence algorithms. Two popular methods for creating accurate ensembles are Bagging and Boosting. These methods rely on "resampling" techniques to obtain different training sets for each of the classifiers, a comprehensive evaluation of both Bagging and Boosting using two basic classification methods (Decision trees and neural networks). Bagging is probably appropriate for most problems, but boosting (either Arcing or Ada) may produce more significant gains inaccuracy [9]. A powerful solution of ensemble machine learning is for an efficient FPGA implemented using Long Short-Term Memory Networks (LSTM) in a classification problem with three different classifiers representing the Base Learners aggregated to obtain the best classifier [10].

The concept of computers is access to data and learning specific tasks without explicitly planned in overcoming problems such as detecting, classifying, diagnosing, selecting, and predicting for all the attributes present in the geological model. When retrieving data and providing the final decision of the analysis probability model produced, it can capture the uncertainty between cause-effect scenarios in the oil and gas industry [11]. Machine learning produces a model with high prediction accuracy. It uses to rank and select to minimize predicted mistakes obtained from the training dataset [12].

This research used well-known machine learning algorithms, such as Naïve Bayes, SVM, and Decision Tree, to classify lithofacies from well log data. Their predictions can be either numerically based on the confusion matrix (also called contingency tables) [13] and by a black-box approach based on the traditional method from an expert's knowledge base on the visualization approach. The motivation research employs Naïve Bayes, SVM, and Decision Tree in supervised and semi-supervised learning only. These three algorithms are not in much use in lithofacies prediction and classification research. This research fundamentally compares these machine learning algorithms to analyze the results directly in the well log data to work on the prediction and classification of lithofacies.

A. SVM Classifier

SVM represents algorithms machine learning-based supervision to perform classification and regression [14], [15]. SVM classifiers and classification are double the maximum margins used to classify the dataset separated linear and non-linear [16]. The ability of SVM to model complex data at linear decision boundaries is very accurate because it functions as a decision-maker [17]. SVM equations use vector support machines to obtain optimal hyperplane for linearly-separable patterns, can expand a pattern that cannot be linearly separated by the transformation of the original data mapped to a new space. Vector Support is the data point closest to the decision-making surface (or hyperplane), as shown in Fig. 1.

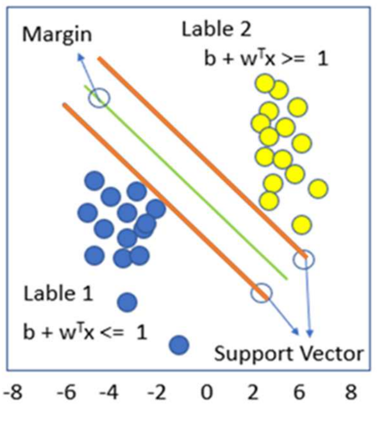


Fig. 1 The data points are the hardest classified, so it requires the support-vector of data that has a direct link to the optimal location of the surface as decision making can formulate as Equation (1).

$$f(y) = w0 \sum_{t=1}^{m} wiXi \qquad (1)$$

With,
W^T = vector (W_1, W_2, W_2,......... W_i)
b = Bias (W0)
X = variable.

The SVM performs well in prediction by using petrophysical logs and inverted seismic attributes data [18].

B. Naive Bayes Classifier

It is a very well-known statistical learning algorithm and is widely recommended as an introductory level classification compared to other algorithms [19]. It is called Naive Bayes or idiots Bayes because it performs probability calculations for each of the hypotheses are simplify in making calculations that can be applied. Naive Bayes predicts a conditional class's probability if a given class is an independent input to each other. This assumption generates the product's discrimination function from the probability that the correct class has given input. Issue predictive modeling classification can frame the calculation of the conditional probability of class labels given sample data.

The machine learning algorithm or model is a specific way of thinking about structured relationships in data. In this way, a model can regard as a hypothesis of data relationships, the relationship between input (X) and output (as in Equation 2). Bayes Theorem is a useful tool in machine learning that provides a way of thinking about the relationship between data and models. Posterior theorem probability P (C | X) can calculate from P (C), P (X), and P (X | C). Therefore,

$$P(C|X) = \frac{P(C)P(X|C)}{\mathrm{P(X)}} \qquad (2)$$

With:
P (C | X) = Class probability target posterior.
P (X | C) = Class probability predictor.
P (C) = Class probability being true.
P (X) = The main probability of the Predictor (y).

C. Decision Tree

Decision Tree is a famous machine learning algorithm used to solve classification problems. The primary purpose of using the decision tree is to predict the target class using the decision rules taken from the previous data [20], generally used in data mining. The purpose is to create a model that predicts the value of the target based on multiple input variables. Use nodes and segments for predictions and classification. The root node classifies instances with different features. The root node can have two or more branches, while leaf nodes represent the classification. The decision tree selects each vertex by evaluating the highest information acquisition among all attributes [21]. The Decision tree describes a combination of mathematical and computational techniques to help with descriptions, categorization, and generalization of the given data set following Equation (3).

$$(x, Y) = (x1, x2, x3 \dots \dots . xk, Y) \qquad (3)$$

The dependent, Y variable, is the target variable that can understand, classify, or describe. Vector x consists of features, x1, x2, x3,......... xk, which uses in the task.

II. Materials and Method

This research used Geolog software to run Python programs and show the interpreter's panels to do analysis easily and check. The interpreter determined the machine learning algorithms if the lithofacies classifications are perfectly matching with the traditional interpretation or not. The following materials and methods can describe as following.

A. Material

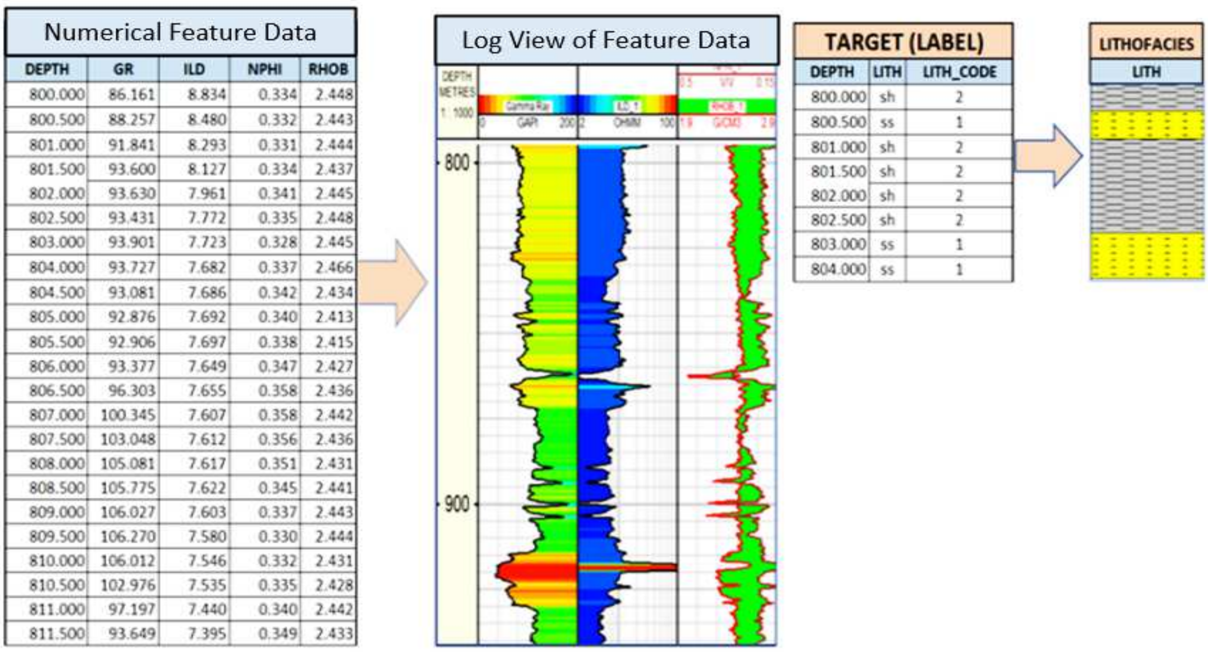
The datasets in this research were obtained from open-source data provided by the American Department of Energy and RMOTC. Only selected well log data are used as they consider representing the oil and gas field of Teapot Dome. Some well log data are used as features. It has been interpreted as a target of lithofacies by an expert for supervised and semi-supervised learning. The process of selecting features in the data contributed to the greatest of the predicted variables (labels) or outputs. Having irrelevant features in data can degrade the model's accuracy, primarily linear algorithms such as linear regression and logistics. The 3 (three) Benefits of making feature selection before modeled data are:

- Reduce Overfitting: Make sure there are not enough data to produce decisions that are influenced by the presence of noise.
- Improve accuracy: Avoid data that can be misleading in order to make the modeling accuracy higher.
- Reduce training time: fewer data, then the algorithm can be trained faster.

The well log data features selected in this research is the log data as follows:

- Gamma-ray (GR) is a method to measure the radiation of the Gamma rays produced by radioactive elements (Uranium, Thorium, Potassium, Radium) in rocks are generally many contained in the shale and fewer contained in sandstone, limestone, dolomite, coal, gypsum. Furthermore, the shale provided a very high response gamma-ray value compared to other rocks with a log analysis of gamma-ray to identify lithology, distinguishing the reservoir zone from a non-reservoir zone.
- Resistivity (ILD) is a method to measure the resistance of reservoir rocks/formations and substances in or around drill holes against electric current expressed in OHM's law. Resistivity log analysis can identify porous and permeable rocks that contain hydrocarbon fluid or water.
- Density (RHOB) is a method to measure the porosity of the formation rocks radiates gamma rays into the drill holes used to find the minerals on the evaporite precipitate, detecting the gas-containing coating, as well as determining hydrocarbons density contained in rock pores. Based on density value, it helps find the rock layer with fluid in the form of gas (low-density value), water fluid, or oil fluid (high-density value).
- Neutron (NPHI) is a method for measuring the hydrogen index (ratio of hydrogen/cm cubic atomic concentration to pure water content at 75 ºF) in rock formations. By measuring the hydrogen index on the pores of the rocks, if the thickness rocks are porous, then the more hydrogen content and the higher the hydrogen index, many rocks containing hydrogen can interpret as having high porosity.

While the given target or label of lithofacies is divided into 2 (two) facies as binary classification, they can be called SS and SH and created by an expert, the interpretation based on the value of the well log data. All features, labels, or targets of lithofacies are shows in the log panels where the numerical values of data can be visualized for an expert to interpret quickly and easily, as shown in Fig. 2.


Numerical Feature Data

DEPTH	GR	ILD	NPHI	RHOB
800.000	86.161	8.834	0.334	2.448
800.500	88.257	8.480	0.332	2.443
801.000	91.841	8.293	0.331	2.444
801.500	93.600	8.127	0.334	2.437
802.000	93.630	7.961	0.341	2.445
802.500	93.431	7.772	0.335	2.448
803.000	93.901	7.723	0.328	2.445
804.000	93.727	7.682	0.337	2.466
804.500	93.081	7.686	0.342	2.434
805.000	92.876	7.692	0.340	2.413
805.500	92.906	7.697	0.338	2.415
806.000	93.377	7.649	0.347	2.427
806.500	96.303	7.655	0.358	2.436
807.000	100.345	7.607	0.358	2.442
807.500	103.048	7.612	0.356	2.436
808.000	105.081	7.617	0.351	2.431
808.500	105.775	7.622	0.345	2.441
809.000	106.027	7.603	0.337	2.443
809.500	106.270	7.580	0.330	2.444
810.000	106.012	7.546	0.332	2.431
810.500	102.976	7.535	0.335	2.428
811.000	97.197	7.440	0.340	2.442
811.500	93.649	7.395	0.349	2.433

TARGET (LABEL)

DEPTH	LITH	LITH_CODE
800.000	sh	2
800.500	ss	1
801.000	sh	2
801.500	sh	2
802.000	sh	2
802.500	sh	2
803.000	ss	1
804.000	ss	1

Fig. 2 From left to right panels, the first panel is the numerical data. The second panel is log curves, representing the numerical values, and the third panel is for target or labels. The fourth panel is for representing the target lithofacies.

B. Method

In this research, well log data in the form of gamma-ray, resistivity, neutrality, and density logs are collected and selected for data processing, transformation, data mining first. The machine learning (Naïve Bayes, SVM, and Decision Tree) can find the pattern or pattern classifications of lithofacies in supervised and semi-supervised to create a model with conditions requiring the change of data and the corresponding requirements of the user's interpretation, as shown in Fig. 3.

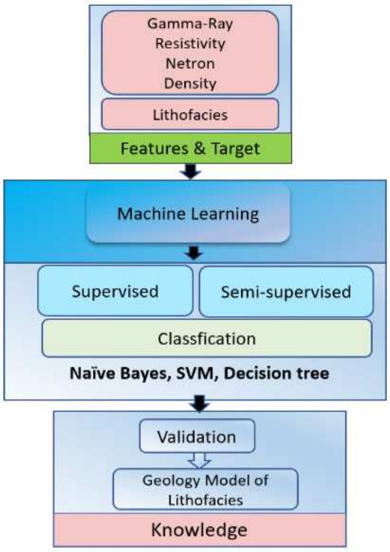


Fig. 3 The workflow of machine learning using Naïve Bayes, SVM, and Decision tree that used in this research.

We used gamma-ray, resistivity, neutron, and density logs data for feature data. And lithofacies as the target. They need to be confirmed and validated their accuracy of lithofacies prediction for output as knowledge discovery. The log interpretation technique uses log curves data to classify sandstone or limestone and hydrocarbon zone lithology, as shown in Fig. 4.

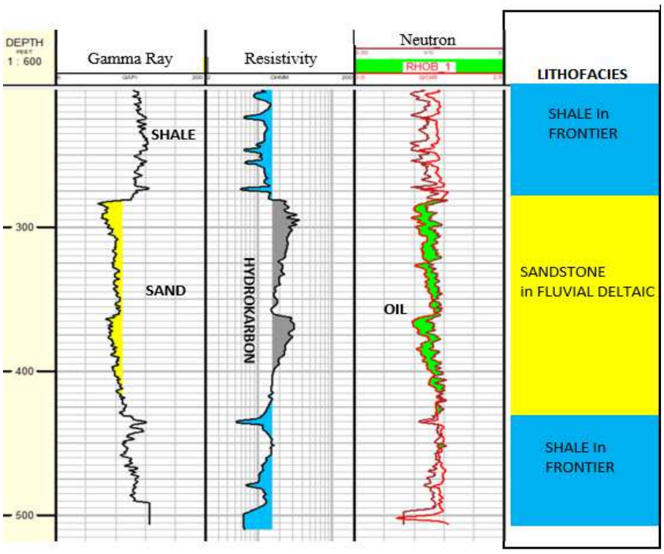


Fig. 4 For oil reservoir zone, gamma-ray and resistivity log, whether deflection goes to the right (high value indicates shale) or left (low value indicating shale).

In the density and neutron log plot techniques, deflection can go to the right or left with arch curve patterns intersecting to show the reservoir zone. If the curves do not cut each other for non-reservoir zones, Neutron-density plots can also show the type of fluid (oil vs. gas vs. water). For the oil reservoir zone, gamma-ray and density-neutron log curves deflection go to the left, and the only resistivity goes to the right. A low value of Gamma-ray is due to the low radioactivity of Th, K, and U contents. Resistivity log responds to hydrocarbons, which are not conducive, providing higher resistivity. The Combination Neutron-Density Log is a combination porosity log. Besides its use as a porosity device, it is also used to determine lithology and detect gas-bearing zones.

This study used 70/30 slicing data. For model evaluation, we compared the accuracy of machine learning algorithms such as SVM, Naïve Bayes, Decision Tree in classifying lithofacies by their root mean square error RMSE value (Equation 4). If predicted machine learning produced a lower RSME value, it implies excellent predictive accuracy.

The following is a formula of RMSE:

$$\mathrm{RMSE} = \mathrm{sqrt}\left(\frac{1}{N}\sum_{i=1}^{N}(predicted_i - actual_i)^2\right) \quad (4)$$

Where: N = amount of data

We also used the black-box approach, which is to get answers by treating the system precisely. This black box does not know its internals, thus understanding its behavior only by processing its input and figuring out what is exciting in the outputs by checking its visualization of classification inside the gamma-ray log panel. This approach can provide an excellent framework to get many insights and has worked well in other fields, all the way from the sciences to the social domain.

III. Results and Discussion

This research aims to evaluate supervised and semi-supervised classification between machine learning algorithms on their accuracy and visualization. After comparing how much a particular model can distinguish all classes, the supervised learning result is shown in Table 1.

TABLE I
CONFUSION MATRIX SUPERVISED

Algorithms	Accuracy	F1	Precision	Recall	Support
Naïve Bayes	0.91	0.90	0.90	0.91	2892
SVM	0.93	0.93	0.93	0.83	2892
Decision Tree	0.99	0.99	0.99	0.99	2892

As in Table 1, the Decision Tree has the best accuracy achieved among the predictors. Moreover, algorithms provide useful predictive in classifications based on the target given and can predict wellbore to make supervised learning predictions. Decision tree and SVM can predict well in supervised learning. Furthermore, all machine learning algorithms offer predictive results or classifications; in sequence, the supervised learning predictions of SVM, Naïve Bayes, and Decision Tree are displayed inside the log curve of the gamma-ray log panel in Fig. 5, 6, and 7, respectively.

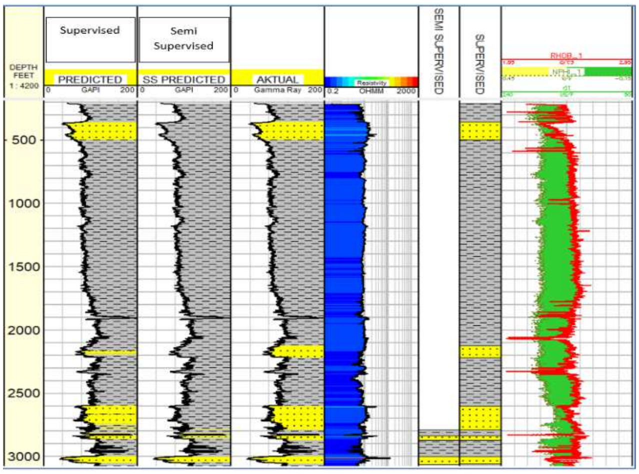


Fig. 5 Classification of SVM semi-supervised and supervised.

The supervised panel shows that the classification of SVM has good predictions in some places along the wellbore and able to meet the criteria of the log interpretation. The accuracy has a high value of 0.93 in supervised and semi-supervised. It has an accuracy of 0.95, and the visualization looks good by an expert for making lithofacies prediction.

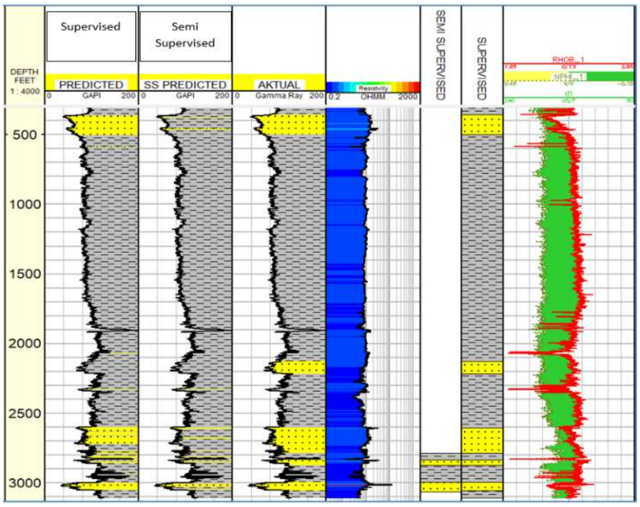


Fig. 6 Classification of Naïve Bayes semi-supervised and supervised.

The supervised. The panel shows that the classification of Naïve Bayes has the right prediction and satisfies the criteria of the log interpretation with an accuracy of 0.91. In semi-supervised panel has an interesting thing, as it is still suitable in performing log prediction even there are no labels. It can predict the target along the well, from start to end depth of well with 0.89 accuracies.

Figure 7 describes that the supervised panel shows that the Decision Tree has advantages in predicting and satisfies the log interpretation criteria with an accuracy of 0.99. In semi-supervised, it has good accuracy of 0.99 as well, but not in doing proper classification along well path data from start depth to the bottom of the well if there is no label.

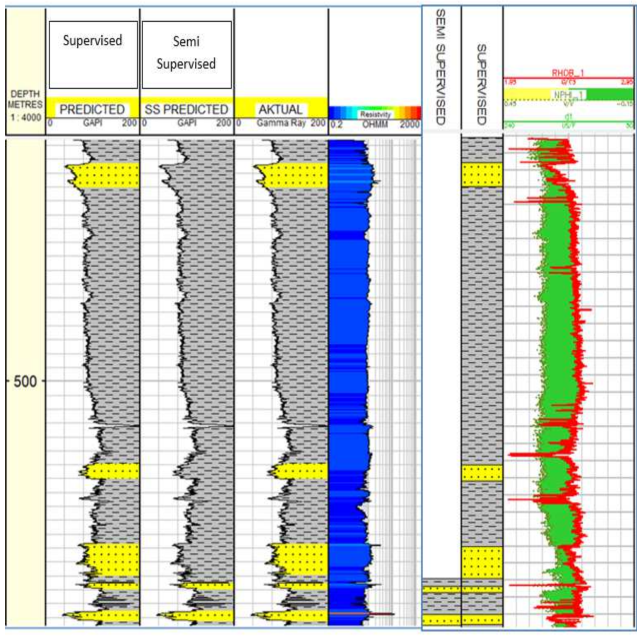


Fig. 7 The result of the classification of semi-supervised and supervised using Decision Tree displayed inside the gamma-ray log curve.

The predicted value from all supervised algorithms machine learning is displayed along with the log curve, as shown in Fig. 8.

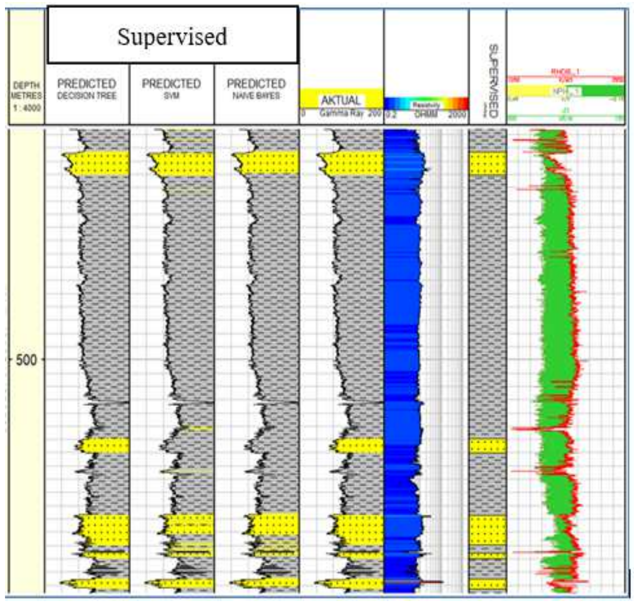


Fig. 8 The result of the classification of supervised machine learning is displayed inside the gamma-ray log curve.

The panel shows that the Decision Tree has advantages in predicting by looking its result in the panel from start depth to the end of the well while Naïve Bayes Tree and SVM are not properly classified. The accuracy of the classification of semi-supervised learning with several machine learning methods, as shown in Table 2.

TABLE II
CONFUSION MATRIX SEMI-SUPERVISED

Algorithms	Accuracy	F1	Precision	Recall	Support
Naïve Bayes	0.89	0.88	0.87	0.89	1933
SVM	0.95	0.95	0.95	0.95	1933
Decision Tree	0.99	0.99	0.99	0.99	1933

As in Table 2, the Decision Tree has the best accuracy achieved among the predictors. Moreover, this is because algorithms provide useful predictive in classifications based on the target but not if there is no target given along wellbore to make semi-supervised learning predictions compared to Naïve Bayes. The decision tree and SVM failed to predict all data without supervision, as shown in the log curve of the gamma-ray panel Fig. 9.

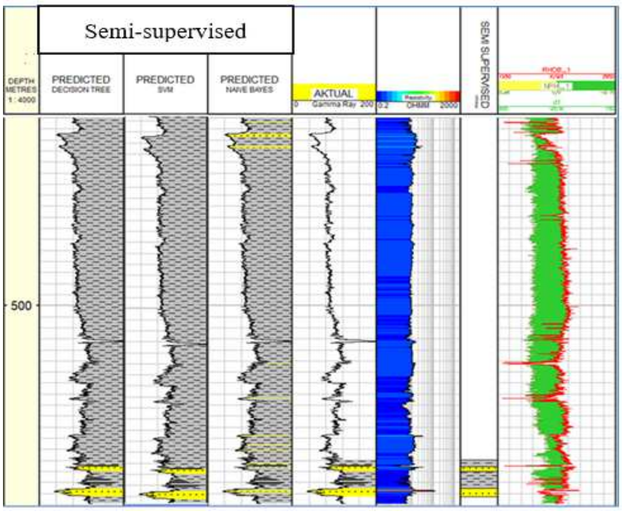


Fig. 9 Classification of semi-supervised machine learning displayed inside the gamma-ray log curve.

The panel shows that Naïve Bayes has advantages in predicting by looking its result in the panel. It can classify from start depth to the end of the well while Decision Tree and SVM are not classifying if there are no labels given. The RMSE calculation is the average value of the number of errors from the three machine learning algorithms for classification of lithofacies model, as shown in Table 3.

TABLE III
RMSE RESULTS CLASSIFICATION OF MACHINE LEARNING

NO	Well	SVM		Naïve Bayes		Decision Tree	
		Accuracy	Train Score (RMSE)	Accuracy	Train Score (RMSE)	Accuracy	Train Score (RMSE)
1	49025104270000	0.95700	0.17	0.90187	0.25	0.95645	0.16
2	49025104280000	0.95690	0.16	0.90575	0.24	0.97011	0.13
3	49025107290000	0.92461	0.18	0.69981	0.48	0.93710	0.19
4	49025107310000	0.99320	0.08	0.97572	0.16	0.99029	0.10
	Average	0.95792	0.1475	0.87078	0.28250	0.963487	0.145

As can be seen in Table 3, low RSME value indicates the variation generated by a prediction approximates the variation of real or observational value. The model train score showed that the Decision Tree is better than Naïve Bayes and SVM.

IV. CONCLUSION

From this research, we concluded that SVM and Decision tree algorithms are superior compare to Naïve Bayes in supervised learning to classifying lithofacies with high accuracy value. However, the best RSME score falls on the Decision tree. As for semi-supervised only naïve Bayes algorithm is superior compared to other algorithms in conducting visual classification. It can make predictions along the wellbore trajectory but not with its accuracy and RSME values than SVM and Decision Tree.

Naïve Bayes wears all predictors to use Bayes rules and assumptions of independence between predictors. However, the Decision Tree wears all predictors assuming dependency between predictors, making it possible for Naïve Bayes to performed predictions along the wellbore trajectory in supervised learning.

SVM itself is like having an advantage in the predictions of lithofacies in supervised and semi-supervised based on the targets given even though the outcome of the accuracy value and RMSE and the visual interpretation is less satisfactory than the Decision tree and Naïve Bayes. This classification technique applied to log data using a machine learning approach such as Naïve Bayes, SVM, and Decision Tree algorithms demonstrate its advantages in lithofacies interpretation workflow for reservoir evaluation as in traditional process. To improve the classification performance, we suggest using an ensemble classifier that consists of combining some machine learning algorithms.

ACKNOWLEDGMENT

We want to thank Emerson for Geolog Software, and Public data provided by Rocky Mountain Oilfield Testing Center (RMOTC) obtained from the following website: http://wiki.seg.org/wiki/Open_dataCenter for Geological Survey management.

REFERENCES

[1] P. Dell'Aversana, "Comparison of different machine learning algorithms for lithofacies classification from well logs," Bollettino Di Geofisica Teorica Ed Applicata, https://doi.org/10.4430/bgta0256, 2019

[2] M.K.Dubois, A.P.Byrnes, G.C.Bohling, S.C.Seals and J.H.Doveton, "Statistically-based lithofacies predictions for 3-D reservoir modelling: Examples from the Panama (Council Grove) Field, Hugoton embayment, Southwest Kansas," American Association of Petroleum Geologists, Annual Convention, Salt Lake City, Utah, 2003.

[3] M.K.Dubois, A.P.Byrnes, G.C.Bohling, S.C.Seals and J.H.Doveton, "Multiscale geologic and petrophysical modelling of the giant Hugoton Gas Field (Permian)," Kansas and Oklahoma, U.S.A. In AAPG Memoir. https://doi.org/10.1306/1215881M883274, 2006.

[4] J. C.Ordóñez-Calderón and S. Gelcich, "Machine learning strategies for classification and prediction of alteration facies: Examples from the Rosemont Cu-Mo-Ag skarn deposit, SE Tucson Arizona," Journal of Geochemical Exploration. https://doi.org/10.1016/j.gexplo.2018.07.020, 2018.

[5] S.Bhattacharya and S.Mishra, "Applications of machine learning for facies and fracture prediction using Bayesian Network Theory and Random Forest: Case studies from the Appalachian Basin," USA. Journal of Petroleum Science and Engineering. https://doi.org/10.1016/j.petrol.2018.06.075, 2018.

[6] T.Wrona, I.Pan, R.L.Gawthorpe and H.Fossen, "Seismic facies analysis using machine learning. Geophysics," https://doi.org/10.1190/geo2017-0595.1, 2018.

[7] X.Wang, S.Yang, Y.Zhao and Y. Wang, "Lithology identification using an optimized KNN clustering method based on entropy-weighed cosine distance in Mesozoic strata of Gaoqing field," Jiyang depression. Journal of Petroleum Science and Engineering. https://doi.org/10.1016/j.petrol.2018.03.034, 2018.

[8] P.Bestagini, V. Lipari and S. Tubaro, "A machine learning approach to facies classification using well logs," 87th Annual International Meeting, SEG, Expanded Abstracts, 2137–2142, doi: 10.1190/segam2017-17729805.1. 2017.

[9] Opitz, D., Maclin, R. Popular Ensemble Methods: An Empirical Study Journal of Artificial Intelligence Research, 11, pp. 169-198, 1999.
[10] Cardarilli, G.C., Di Nunzio, L., Fazzolari, R., Giardino, D., Matta, M., Re, M., Silvestri, F., Spanò, S. Efficient ensemble machine learning implementation on FPGA using partial reconfiguration Lecture Notes in Electrical Engineering, 550 (9783030119720), pp. 253-259, 2019.
[11] O.Bello, C. Teodoriu, T.Yaqoob , J.Oppelt, J.Holzmann and A.Obiwanne, "Application of artificial intelligence techniques in drilling system design and operations: A state of the art review and future research pathways," Society of Petroleum Engineers - SPE Nigeria Annual International Conference and Exhibition. https://doi.org/10.2118/184320-ms, 2016.
[12] James, G., Witten, D., Hastie, T., & Tibshirani, R," An Introduction to Statistical Learning with Applications in R," In Springer. https://doi.org/10.1016/j.peva.2007.06.006, 2013.
[13] K. M. Ting, "Confusion Matrix. In Encyclopedia of Machine Learning and Data Mining," https://doi.org/10.1007/978-1-4899-7687-1_50, 2017.
[14] L.Hamel, "Knowledge Discovery with Support Vector Machines. In Knowledge Discovery with Support Vector Machines," https://doi.org/10.1002/9780470503065, 2009.
[15] E. Alpaydin, "Introduction to Machine Learning Ethem Alpaydin. Introduction to Machine Learning, "Third Edition, 2014.
[16] J.Han, M. Kamber and J. Pei, "Data Mining: Concepts and Techniques. In Data Mining: Concepts and Techniques," https://doi.org/10.1016/C2009-0-61819-5, 2012.
[17] I. Guyon, "A practical guide to model selection," In: Marie, J. (Ed.), Proceedings of the Machine Learning Summer School. Canberra, Australia, January 26 - February 6, Springer Text in Statistics, Springer p.37, 2009.
[18] M.A. Sebtosheikh and A.Salehi, Lithology prediction by support vector classifiers using inverted seismic attributes data and petrophysical logs as a new approach and investigation of training data set size effect on its performance in a heterogeneous carbonate reservoir, Journal of Petroleum Science and Engineering, 134 143–149, 2015.
[19] Ray, S. "6 Easy Steps to Learn Naive Bayes Algorithm (with code in Python)," Analytics Vidhya, 2015.
[20] P. A. Harrison, R.Dunford, D. N.Barton, E.Kelemen, B.Martín-López, L.Norton, M.Termansen, H. Saarikoski, K.Hendriks, E.Gómez-Baggethun, , B.Czúcz, M.García-Llorente, D.Howard, S. Jacobs, M.Karlsen, L.Kopperoinen, A.Madsen, G.Rusch, M.van Eupen, G. Zulian, "Selecting methods for ecosystem service assessment: A decision tree approach ecosystem services," https://doi.org/10.1016/j.ecoser.2017.09.016, 2018.
[21] A. S.J.Iyer and R.Sumbaly, "Diagnosis of diabetes using classification mining techniques," International Journal of Data Mining & Knowledge Management Process, https://doi.org/10.5121/ijdkp.2015.5101, 2015.