
Lithofacies Classification using Supervised Machine Learning Algorithm

Stefanus Yudi Irwan : yudi.stefanus22@gmail.com

Abstract

Facies classification is still nowadays carried out by experienced geologists or geophysicist by using well logs measurement data such as GR, RHOB, NPHI, etc. Facies classification is carried out by studying the lithological properties of rock by bringing formation samples to the surface which requires coring tools that are expensive. Machine learning classification can be an alternative to classify this rock layer in the reservoir straight from the logging data. The aim of this research is to develop an effective supervised machine learning model to classify rock facies. From the 28 supervised machine learning model trained, comprises of 7 different algorithm and 4 different dataset type, this research found that extreme gradient boosting classifier perform better than the other supervised machine learning algorithm used in this research. This XGBoost Classifier performance value comprises of accuracy 52.7%, adjacent accuracy 86.6%, cv-score 60.8%, and ROC-AUC score of 86.5%. In terms of best performance, this XGB Classifier has a trade-of in the longer training time compare to others supervised machine learning algorithms, this training time could be costly if production scale machine learning system will be deployed

A. Introduction

Gas and oil reservoirs have been the focus of oil and gas well exploration activity in many countries. The attempt is to search for zones with a high volume of hydrocarbon beneath the surface. The aim of this exploration is to bring that trapped hydrocarbon to the surface for people's consumption. As the Earth changes over time, different rock-type layers (or lithofacies) are created. These subsurface layers form the time-based cyclic sequence, where layers that are superimposed on top of others are geologically younger. Certain layers and layer sequences, such as shale, are known to be impermeable to gas and/or liquid. Oil and natural gas then become trapped by these layers, making it possible to drill wells (also called cores) to reach the supply and extract for use. The facies (part of the layer that differs from the adjacent layers by lithological composition) contain characteristic features found in the core of reservoir rock samples that were taken from wells. They are becoming the basis for characterizing and constructing reservoir models. The ability to predict where these layers and known sequences are, based on properties of core samples, provides the potential for more efficient "payzone" drilling. [2]

To date, rock facies (subsurface layer) classification has largely used wireline log measurements together with multivariate statistical methods, fuzzy mathematical approaches, and artificial neural networks. Wireline well

logs are physical and chemical rock measurements recorded by lowering logging tools with specialized sensors into wells after they have been drilled. Properties such as permeability, porosity, and liquid content can be extracted using specific measurement mediums. Electrical measurements offer information on saturation (e.g., oil or water); while acoustic measurements can provide information about grain size. Thus, certain rocks and layers will exhibit characteristic measurement signatures based on what they contain. Experts like geologists or petrophysicist then analyze these data to determine the rock type and sequence [2]

Conventionally, facies definition and distribution are very time-consuming and expensive since it involves the in-depth analysis of well-cores, well-logs, downhole measurements, analogue studies, and is holistic in nature. Such manual analysis of large datasets by a reservoir geologist or geo-physicist must be subjective to some extent, leading to increased uncertainties. [4]

In recent years, there has been an increasing demand for data science across many disciplines. This is because technological advances have improved data storage capabilities as well as how data can be obtained (e.g., real-time data). Manually interpreting data that are exponentially growing in volume has obvious management and analysis challenges. Machine learning automates pattern recognition of big data in an efficient manner. [3] Machine learning processes provide the opportunity to make the process of facies recognition more

automatic and objective. The accuracy of the result depends on the relevance and complexity of the method, and critically on the relevance and quality of the attributes that represent the input physical properties.

In machine learning, facies recognition may be approached by using either clustering or classification. Table 1 shows the differences between these two approaches. Clustering is simpler, does not require foreknowledge of the number or properties of the facies, and does not need training. However, it can only separate facies on the basis of the natural similarity or difference in the properties (which are called attributes in machine learning parlance) and hence it provides more general solutions. These aspects of clustering give it advantages in facies recognition where only an attribute dataset exists. However, if facies in one well have already been defined and it is necessary to find to which of these facies a sample or samples in another well belong, classification must be used.

Table 1
Differences and similarities between Clustering and Classification.

Clustering	Classification
Less complex	More complex
Unsupervised	Supervised or unsupervised
Facies clustered into groups according to similarities in their natural properties (attributes)	Facies classified into known and predetermined classes on the basis of their natural properties (attributes)
Does not need facies (classes) to be already defined.	Needs facies (classes) to be already defined.
Method needs to be tuned for optimum efficacy	Method needs to be tuned for optimum efficacy
No training needed	Method needs to be trained on known facies of the same type.
Validation testing is required.	Validation testing is required.
Can only be used to recognise general groupings such as facies.	Can be used to classify data by finer distinctions in their attributes.

Classification in machine learning can be achieved through supervised machine learning. This supervised technique used labelled data to understand patterns and then label other data based on the trained patterns. There are a lot of supervised algorithms for classification, like K-Nearest Neighbor, Support Vector Machines, Logistic Regression, Decision Trees, Gradient Boosting, etc. The aim of this study is to find which algorithm will provide the best lithofacies prediction of the test data.

This is how the rest of the paper is organized. Section 2 will discuss the related work of classifying lithofacies using machine learning algorithm. Section 3 will discuss about the dataset that is used for this research. Section 4 will discuss the method of this research. Section 5 will discuss the result, and then the conclusion in section 6.

1. Related Works

In the petroleum industry, many machine learning-based techniques have recently emerged to increase the efficiency and accuracy of lithofacies classification. For example Yadigar Immamverdiyev and Lyudmila Sukhostat [1] construct a model to interpret the well-log data comprises of the photoelectric effect (PE), gamma ray

(GR), resistivity logging (RL), neutron-density porosity difference (DPHI), and average neutron density porosity (PHIA) by using a model architecture based on a one-dimensional convolutional neural network (1D-CNN) for facies classification. The model was trained on various optimization methods, such as Adagrad, Adadelata, and Adamax, and the result was compared with another research about RNN and LSTM models. The result shown that the best model to predict the lithofacies was achieved by using 1D-CNN with Adagrad Optimization, and the other 1D-CNN model (Adadelata and Adamax) shown that it has the accuracy better than the others model such as SVM, KNN, RNN, and LSTM.

Christopher M. Gifford and Arvin Agah [2] using Multi-Agent Collaborative Learning Architecture. The data was obtained from The Kansas Geological Survey (KGS). The research combine several algorithms such as Naïve Bayes, Decision Tree, Instance-Based KNN, Neural Network, etc to be several groups and compare the testing accuracy. The result was shown that by utilizing Multi-Agent Collaborative Learning Architecture, the researchers was able to obtain approximately 84.5% accuracy from data test, and this is improved by 6.5% from the Kansas Geological Survey (KGS) research.

Dunham, et all [3] make a model to classify the lithofacies by using a Semi-Supervised Gaussian Mixture and compare the result with XGBoost Classifier. They also use data from Kansas Geological Survey, provided by a Lithofacies competition in 2016. The result was shown that the developed Semi-Supervise Gaussian Mixture was able to out-perform the XGB accuracy by approximately 5%. ssGM also provide a benefit as a simple semisupervised algorithm with only two hyper-parameter than a complex supervised algorithm such as XGBoost.

Glover et all [4] use a large database of high-quality poroperm, electrical, mercury injection capillary pressure and nuclear magnetic resonance spectroscopy measurement. This research use un-supervised learning algorithm to classifies lithofacies of oil and gas well. For a 3 cluster problem, the research found that the Expectation Maximisation (92.57% success) and two types of Kmeans approaches (89.60% and 91.09%) provide the best methods.

Antariksa et all [5] conduct research in Indonesia that took place in the Tarakan Basin. They used two separate well log datasets from the Tarakan Basin. The result showed that Random Forest and Gradient Boosting outperformed the other model in the experiment with accuracy of 87.49% and 87.01%, respectively.

This research will be focused on exploring supervised learning algorithm and find the best model for the dataset provided by SEG machine learning competition in 2016 that was made public by Kansas Geological Survey (KGS).

B. Dataset

The dataset comprises of ten oil and gas wells, nine are real well and 1 are synthetic well. Nine wells are drilled in the Hugoton and Panoma fields of southwest Kansas and Northwest Oklahoma. All ten wells consist of wireline log data and geologic variables as input for this research. Based on figure 1, for a whole dataset, the ratio for every real well is nearly even, but for the synthetic well (F9) the ratio is small compared to the other because this well is created for compensating the category BS (Phylloid-Algae Bafflestone) that is scarce in others well.

There are nine output labels that this research tries to predict (Table 1). This label consists of 9 types of lithofacies commonly encountered in oil and gas drilled well. The actual condition in the oil well, the rock is blended with one another rock at every facies layer. This makes the classification problem becoming harder since in the real condition one rock is able to blend with one another. For compensating the reality this research will use the term “Adjacent Facies” to denote the layer. The adjacent facies state the proximity of one facies to another facies. Hence at prediction process we state that the facies and its adjacent is one label. Next in this research accuracy calculation will be counted based on the prediction of lithofacies and the prediction of adjacent lithofacies.

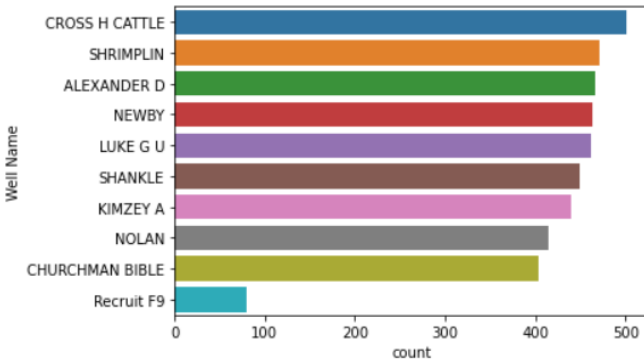


Figure 1. Data point for every well

Facies	Adjacent Facies	Classes of Rocks
SS	CSiS	Sandstone
CSiS	SS, FSiS	Coarse Siltstone
FSiS	CSiS	Fine Siltstone
SiSh	MS	Marine Siltstone and Shale
MS	SiSh, WS	Mudstone (limestone)
WS	MS, D	Wackestone (limestone)
D	WS, PS	Dolomite
PS	WS, D, BS	Packstone-grainstone (limestone)
BS	D, PS	Phylloid-algal bafflestone (limestone)

Table 1. List of Lithofacies

As an input of this research, the dataset provides 7 feature variables, 5 features from wireline measurement, and 2 features from geological variables. Gamma Ray

measures the gamma-ray activity of the lithofacies, this measurement will generate different values for different lithofacies. The resistivity log measures the lithofacies through the injection of an electromagnetic field into the formation, the feedback signal is then interpreted as the resistivity value for the formation. Usually, different lithofacies will generate different resistivity. Radioactive-based measurements like Photoelectric Effect, Neutron-Density Porosity Difference, and Average Neutron Density Porosity use radioactive sources to measure the formation in the borehole, radioactive feedback signal then could be interpreted as the Porosity of the formation and Photoelectric effect. This also usually generates a different value for different lithofacies.

Non-Marine Marine indicator is a categorical variable that states the lithofacies deposition in marine or non-marine environments. In exploratory Data Analysis activity, it is found that from the data the majority of classes CSiS, FSiS, and SS is belongs to Non-Marine Indicator, whereas the rest of the class belongs to Marine indicator. This indicator was achieved through geological study by some experts. The relative position doesn't necessarily contribute to the facies classification, this data is explained about the relative position of one labeled lithofacies in the formation layer.

Features	Explanation	Measurement Type
GR	Gamma Ray	Wireline
ILD_log10	Resistivity_log	Wireline
PE	Photoelectric Effect	Wireline
DelthaPHI	Neutron-Density Porosity Difference	Wireline
PHIND	Average Neutron-Density Porosity	Wireline
NM_M	Non-Marine Marine Indicator	Geologic Constraining Variables
RELPOS	Relative Position	Geologic Constraining Variables

Table 2. List of Input Features

C. Methods

C.1. Data Pre-processing

Raw data is comprised of 10 CSV files that represent the well measurement from 10 different well. Data were combined into a single Pandas data Frame, comprises of 11 features and 4149 rows. This research use ‘CHURCHMAN BIBLE’ well data as test data for comparison between labelled facies and predicted facies. The rest of the 9 well data was used as train data for the model. There are some missing values in the train data and test data for PE measurement. This null data in train data is then imputed by using the average value of PE measurement for every label. The null data in test data is imputed by the average value in train data to avoid data leakage.

After the null value was handled, the data pre-processing phase continued to check the data type for

every column. Figure 2. Show the data type for every column used in this research

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3745 entries, 0 to 4148
Data columns (total 11 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Facies       3745 non-null   object
1   Formation    3745 non-null   object
2   Well Name    3745 non-null   object
3   Depth        3745 non-null   float64
4   GR           3745 non-null   float64
5   ILD_log10    3745 non-null   float64
6   DeltaPHI     3745 non-null   float64
7   PHIND        3745 non-null   float64
8   PE           3745 non-null   float64
9   NM_M         3745 non-null   int64
10  RELPOS       3745 non-null   float64
dtypes: float64(7), int64(1), object(3)
memory usage: 480.1+ KB
```

Figure 2. Train Data Info

Figure 2 shows that in this research we have 4 categorical variables and 7 numerical variables. Although NM_M data type is 'int64' this data represents the category of marine and non-marine facies, hence we state it as the categorical variable. Complete data pre-processing phase is depicted in figure 3.

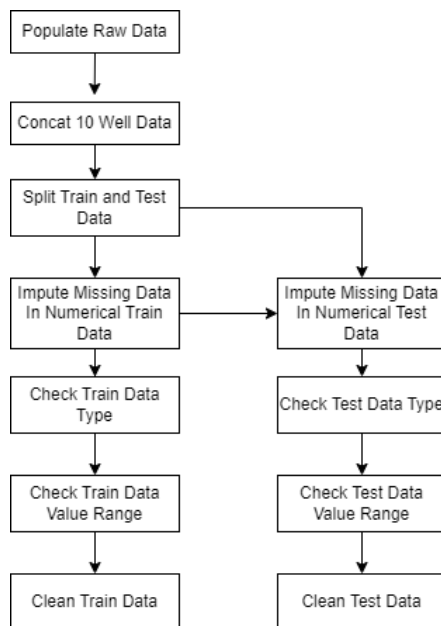


Figure 3. Data Preprocessing Block Diagram

C.2. Feature Engineering

After clean data was formed from the data pre-processing phase, the next phase is to form appropriate data for training machine learning models. From the Data analysis phase, this research won't use several features such as Formation, Well Name, Depth, and RELPOS because these features don't contribute to the classification. Formation and Well name is just a nickname data from the geologist. Depth is a unique data that represent the depth of the lithofacies. RELPOS (Relative Position) data is highly correlated to depth data for every formation name. Feature selection then drop this

feature from the clean data to not use this features for modelling.

The data is then split into input numerical data, input categorical data, and output data. Input numerical data comprises GR, ILD_log10, PE, DeltaPhi, and PHIND features. From the data pre-processing phase before this numerical features have already cleaned, so there is no need to handle this feature in feature engineering phase.

Input categorical data is comprises of NM_M data, to handle this categorical data, NM_M will be encoded using One Hot Encoder to alter the data type into numerical data. NM_M feature will be transaled into Marine and Non-Marine feature by this One Hot Encoder, and added to input numerical train data as new numerical features.

After every train data feature becomes numerical data, features are then normalized so it has an average value of 0 and a standard deviation of 1 by Standard Scaller.

Output features on train data are Facies data. This is comprised of facies abbreviation name, to handle this data we use a label encoder to translate this abbreviation name into numerical data. The facies abbreviation name and numeric representation is shown in table 2

Facies	Numeric Representation
SS	0
CSiS	1
FSiS	2
SiSh	3
MS	4
WS	5
D	6
PS	7
BS	8

Table 2. Facies Label Encoding

Because this facies data is an unbalance data, this research will try to balance the train data into random under sampling data, random over sampling data, and smote sampling data. Hence at the end of the feature engineering phase for train data we have 4 set of train data unbalance train data, under-sample train data, over-sample train data, and smote train data, to increase the variety of model. The distribution of category for every train data depicted in figure 4 – figure 7. As you can see from the figure 4 until figure 7, unbalance train data has unbalance output category, whereas other train data has balance train category.

Under-sampling train data take the least amount from the unbalance data, which is Dolomite, and cut other categories data. Hence from under sampling train data we have less data for training purpose. Over sampling and Smote train data take the largest amount from the unbalance train data, which is CSiS, and resample other category randomly so other category has much more data to follows the CSiS data. Hence at the end we have more

data as train data for the model from over sampling and Smote

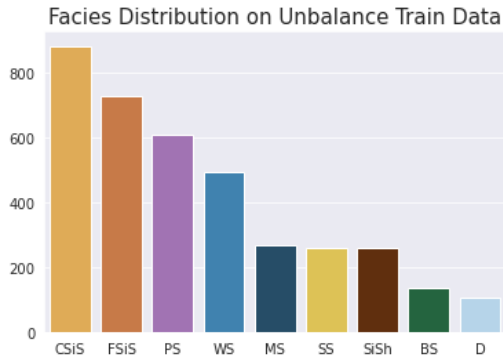


Figure 4. Facies Distribution on Unbalance Train Data

The step that has been done to train data in this feature engineering phase is also applied to test data, without balancing the test data. The recap of this feature engineering phase is depicted in figure 8.

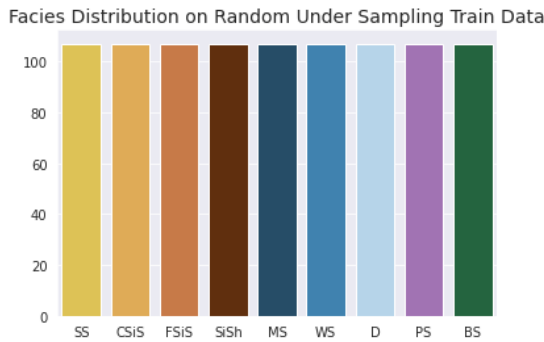


Figure 5. Facies Distribution on Random Under Sampling Train data

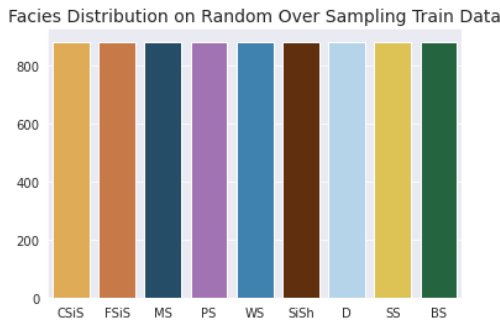


Figure 6. Facies Distribution on Random Over Sampling Train data

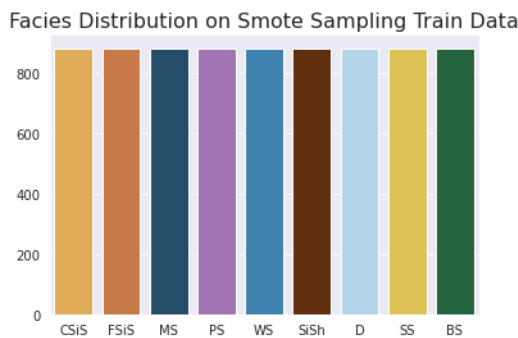


Figure 7. Facies Distribution on Smote Sampling Train data

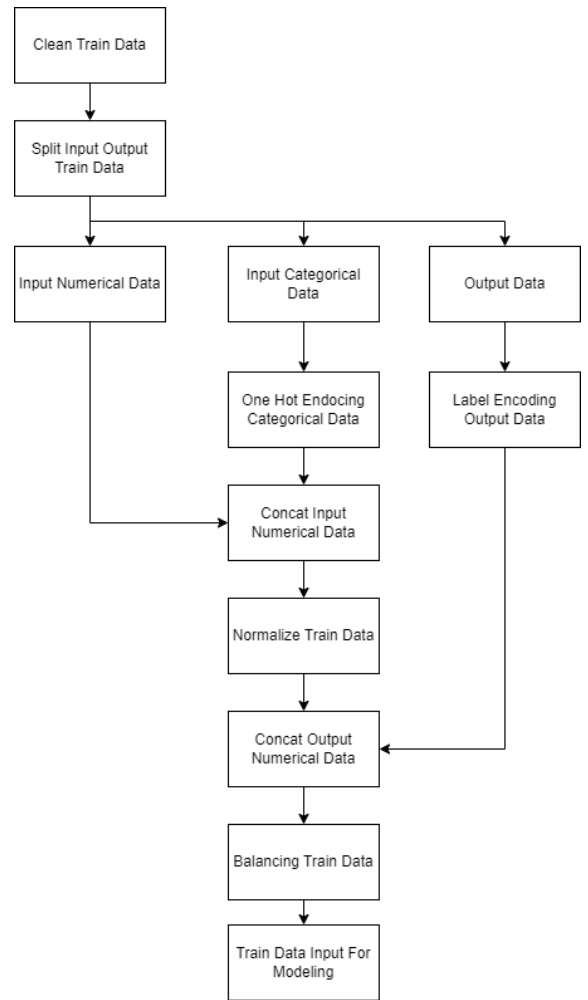


Figure 8. Feature Engineering Block Diagram

C.3. Modeling

This research utilize several machine learning model classifier for facies classification modeling. There are KNeighbor Classifier, Decision Tree Classifier, Bagging Classifier, Logistic Regression Classifier, Support Vector Machine, Random Forest Classifier, and XGBoost Classifier. Modeling was done by using 4 datasets, there are unbalance data, under-sample data, over-sample data, and smote-sample data.

From the 7 machine learning models and 4 datasets, this experiment will produce 28 machine learning models for facies classification, and at the end of the modeling phase, the accuracy of the data test will be assessed to pick one of the best models from the accuracy score. The accuracy score means how well the model can predict the facies from the input data compared to the actual facies from the output data.

Because this research use so many supervised machine learning model, the training phase plus the hyperparameter tuning for every model, will take approximately an hour to complete, this is relatively a long time for the production phase, but in research, this is not the main concern.

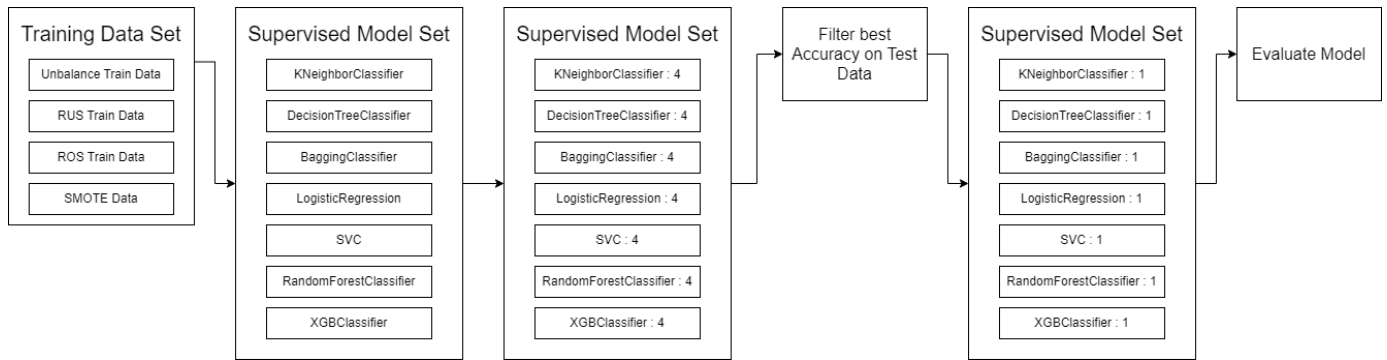


Figure 9. Modeling Facies Classifier

This research use Dummy Classifier model as a Base model trained with unbalance train data. This Dummy Classifier act as a comparator for every machine learning model. The Dummy Classifier accuracy score is 0.24 and the adjacent accuracy score is 0.50. This accuracy score means that Dummy Classifier could predict 24% facies true compare to actual facies label in unbalance train data, and adjacent accuracy score means this model could predict 50% facies and its adjacent facies true compare to actual facies label data in unbalance data. Besides accuracy and adjacent accuracy score, the models were also evaluated using cross-validation score, and ROC-AUC value

C.4. Model Evaluation

1. Evaluation by Accuracy

The result of the evaluation shows that, at well test data (CHURCHMAN BIBLE) all model has accuracy beneath 0.6. This is still acceptable since at the actual facies are blended rock. From horizontal bar graph we can see that XGBClassifier has the highest accuracy on test data for 52.7% and Decision Tree Classifier has the smallest accuracy on test data for 46.5%. The accuracy score on test data for other model depicted in figure 9

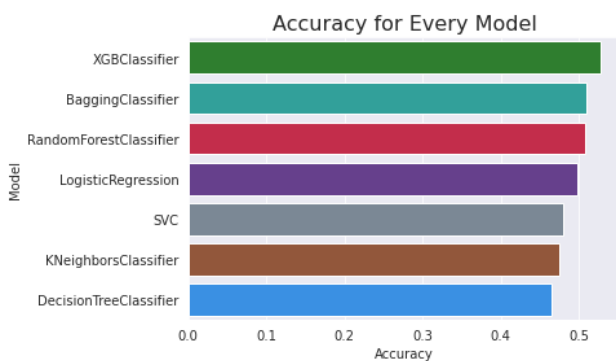


Figure 9. Accuracy on Test Data

2. Evaluation by Adjacent Accuracy Score

Adjacent accuracy on test data has higher value than just accuracy on test data. For all model trained the value is

roughly more than 80% for adjacent facies accuracy. XGB Classifier again has the highest adjacent facies value for 86,8% and again Decision Tree Classifier become the model with the smallest adjacent accuracy on test data for 80,2% The rest of the model adjacent accuracy depicted in figure 10.

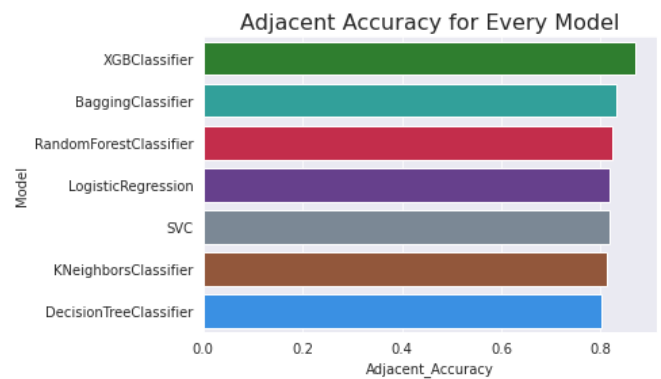


Figure 10. Adjacent Accuracy on Test Data

3. Evaluation by Cross-Validation

Cross-validation score show us that Random Forest Classifier has the highest cv-score for 89.34%, whereas KNN has the smallest cv-score for 49,1%. If we compare this cv score with the accuracy score there are much different in value from Random Forest classifier and Support Vector Machine. This is because this two model overfit in train data, hence when tested on the test data the accuracy drop from its cross validation score. This overfit could mean that there are behavioural difference between the train data and the test data. Unfortunately Random Forest Classifier and Support vector machine couldn't generalize this test data, hence the accuracy is drop from it's cross-validation score.

There are not much different for the rest of the model, this means that the models are well trained by using train data. If we exclude the Random Forest Classifier and Support Vector Machine, XGB Classifier has the highest cv score for 60.85% whereas KNN has the smallest cv score for 49,1%. The rest of the model cross validation score are depicted in figure 11

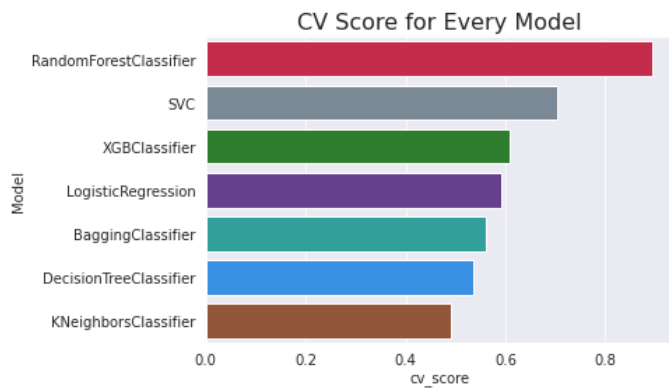


Figure 11. Cross Validation Score for Every Model

4. Evaluation by ROC-AUC Score

For the classification case, we could use the ROC-AUC curve as an evaluation method. Figure 12 shows that XGB Classifier has the highest ROC-AUC score of 86.5% whereas Decision Tree Classifier has the smallest ROC-AUC score of 82%. Overall every model has ROC-AUC score up to 80%.

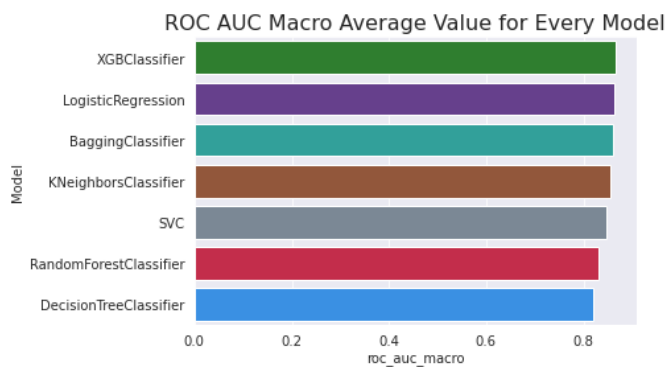


Figure 12. ROC AUC Score for Every Model

One example of ROC-AUC curve was shown in figure 13, as you can see that for multi-class classification ROC-AUC curve was constructed by computing average TPR and FPR for every category, hence this research use micro average ROC-AUC score and macro average ROC-AUC score. For comparison purpose this research use macro average ROC-AUC score.

1. Evaluation by Facies log plot

After every model was trained, we can then compare its performance to predict the CHURCHMAN BIBLE well measurement data. Figure 14 shows that every model could predict the majority of facies when the layer doesn't variate much, like at the depth around 3075 and 2960. But when it comes to variative layer like in the depth around 3050 and 3100 the predicted lithofacies become clearly different with actual facies. XGB Classifier with the best performance evaluated from accuracy, adjacent accuracy, cv-score, and ROC-AUC curve seems could predict the layer better than any other model.

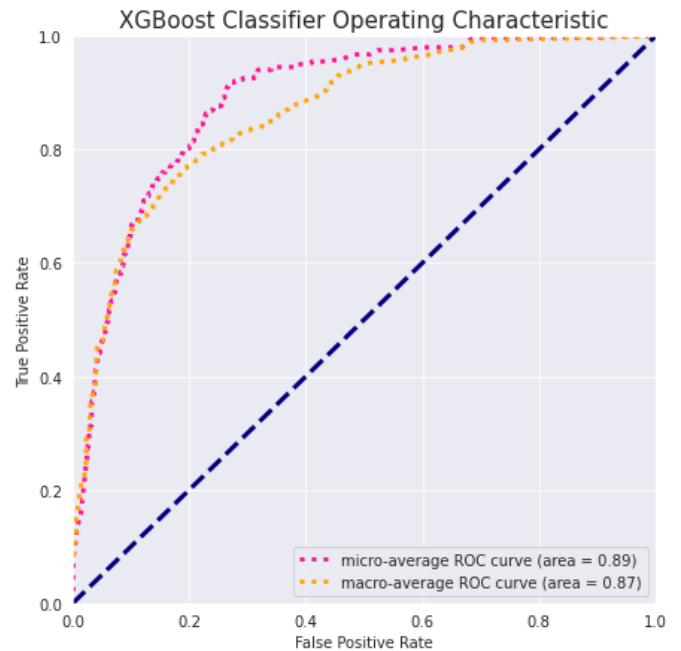


Figure 13. ROC-AUC Curve for XGBoost

Comparing XGB Classifier performance with KNN, we could see that the KNN model with an accuracy of 47.5% has more layers than actual layers for CHURCHMAN BIBLE well

D. Conclusion

This research found that for facies classification case by comparing the performance of the models, the best supervised machine learning algorithm is XGBoost Classifier with an accuracy of 52.7%, adjacent accuracy of 86.6%, cv-score of 60.8%, and ROC-AUC score of 86.5%. One concern about this XGBoost classifier is although the performance score is good compare to others model the training time is longer than other simple supervised machine learning model like K-Neighbor or Decision Tree, this training time could be the consideration for design the facies classifier for production purpose.

E. References

- [1] Imamverdiyev, Y., Sukhostat, L., 2019, Lithological facies classification using deep convolutional neural network. *Journal of Petroleum Science and Engineering* 174 (2019) 216–228
- [2] M. Gifford, C. Agah, A., 2010, Collaborative multi-agent rock facies classification from wireline well log data, *Engineering Applications of Artificial Intelligence* 23 (2010) 1158–1172
- [3] W. Dunham, M. Malcolm, A. Kim Welford, J. 2020, Improved well log classification using semisupervised Gaussian mixture models and a new hyper-parameter selection strategy, *Computers and Geosciences* 140 (2020) 104501
- [4] W.J. Glover P., K. Mohammed-Sajed, O., Akyiiz, C., Lorinczi, P. 2022, Clustering of facies in tight carbonates using machine learning, *Marine and Petroleum Geology* 144 (2022) 105828

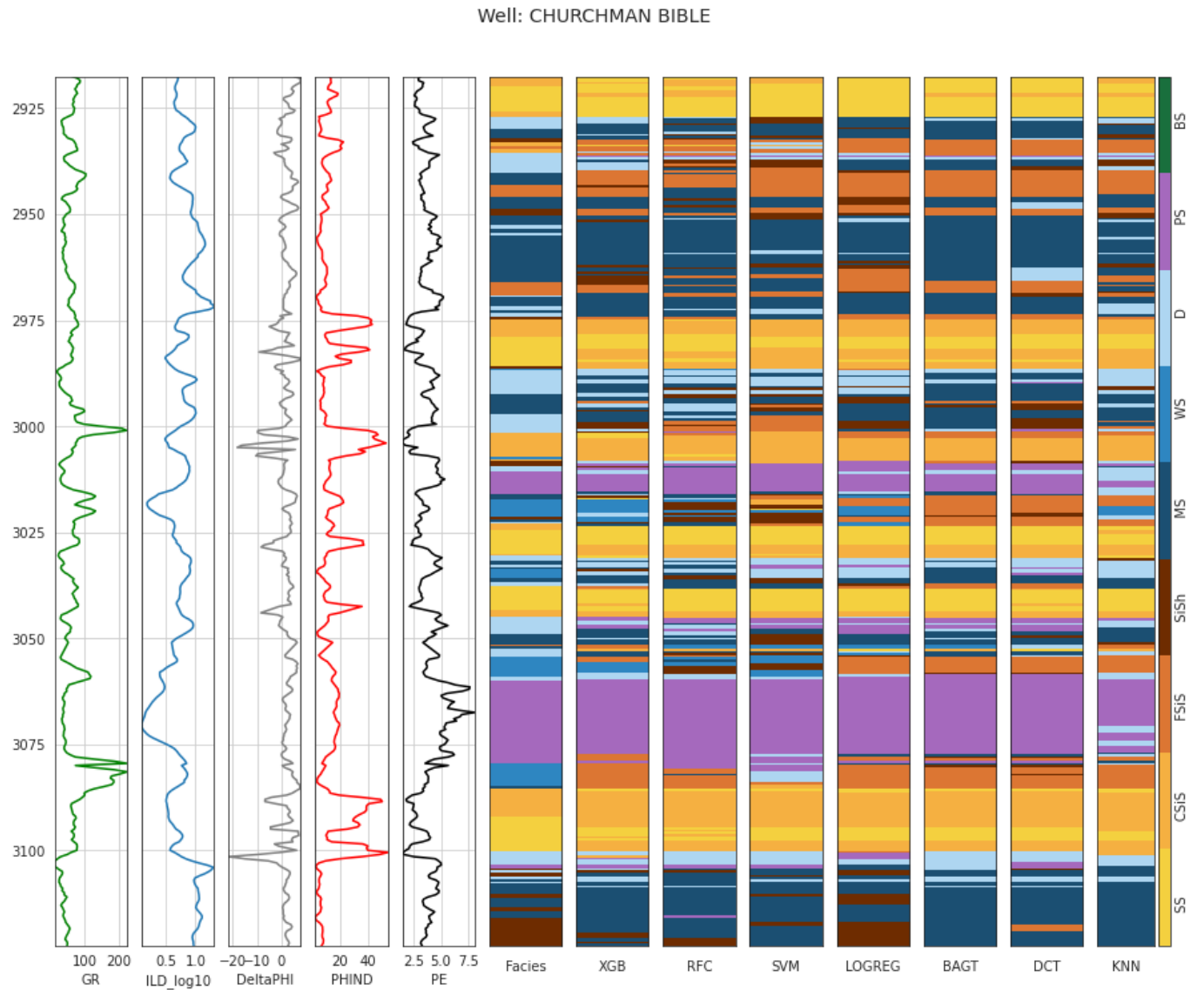


Figure 14. Comparison of Predicted Lithofacies for Every Model