โครงการการเรียนการสอนเพื่อเสริมประสบการณ์

การทำเหมืองข้อมูลของข้อมูลหยั่งธรณีหลุมเจาะ
ด้วยแบบจำลองการตัดสินใจแบบต้นไม้

โดย

นายธันยบูรณ์ สุธาศิริกุล
เลขประจำตัวนิสิต 5832716123

โครงการนี้เป็นส่วนหนึ่งของการศึกษาระดับปริญญาตรี
ภาควิชาธรณีวิทยา คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2561

การทำเหมืองข้อมูลของข้อมูลหยั่งธรณีหลุมเจาะ
ด้วยแบบจำลองการตัดสินใจแบบต้นไม้

นายธันยบูรณ์ สุธาศิริกุล

โครงงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต
ภาควิชาธรณีวิทยา คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2561

DATA MINING OF WELL LOGS
USING DECISION TREE BASED MODEL

MISTER THANYABOON SUDHASIRIKUL

A Project Submitted in Partial Fulfillment of the Requirements
for the Degree of Bachelor of Science Program in Geology
Department of Geology, Faculty of Science, Chulalongkorn University
Academic Year 2019

หัวข้อโครงงาน	การทำเหมือนข้อมูลของข้อมูลหยั่งธรณีหลุมเจาะ ด้วยแบบจำลองการตัดสินใจแบบต้นไม้
โดย	นายธันยบูรณ์ สุธาศิริกุล
สาขาวิชา	ธรณีวิทยา
อาจารย์ที่ปรึกษาโครงงานหลัก	ผู้ช่วยศาสตราจารย์ ดร.วรัญทร คณิตปัญญาเจริญ

วันที่ส่ง 13 พ.ค. 62
วันที่อนุมัติ 13 พ.ค. 62

อาจารย์ที่ปรึกษาโครงงานหลัก
(ผู้ช่วยศาสตราจารย์ ดร.วรัญทร คณิตปัญญาเจริญ)

Project Title	DATA MINING OF WELL LOGS USING DECISION TREE BASED MODEL
By	Mister Thanyaboon Sudhasirikul
Field of Study	Geology
Project Advisor	Assistant Professor Dr.Waruntorn Kanitpanyacharoen

Submitted date....13 May 2019....

Approval date....13 May 2019....

....Waruntorn....

Project Advisor

(Assistant Professor Dr. Waruntorn Kanitpanyacharoen)

ธันยบูรณ์ สุธาศิริกุล : การทำเหมืองข้อมูลของข้อมูลหยั่งธรณีหลุมเจาะด้วยแบบจำลองการตัดสินใจแบบต้นไม้ (DATA MINING OF WELL LOGS USING DECISION TREE BASED MODEL) อ.ที่ปรึกษาโครงงานหลัก : ผู้ช่วยศาสตราจารย์ ดร.วรัญทร คณิตปัญญาเจริญ, 59 หน้า

การหยั่งธรณีหลุมเจาะเป็นการสำรวจค่าธรณีฟิสิกส์ต่าง ๆ ของชั้นหินผ่านหลุมเจาะสำรวจ เพื่อให้ทราบธรณีวิทยาใต้ผิวดิน การแปลความหมายข้อมูลธรณีหยั่งหลุมเจาะเป็นขั้นตอนที่ใช้เวลาและจำเป็นต้องอาศัยประสบการณ์และความชำนาญของนักธรณีวิทยาที่รับผิดชอบ สถิติจึงอาจนำมาวิเคราะห์เพื่อช่วยให้การแปลความหมายมีประสิทธิภาพมากยิ่งขึ้น การศึกษานี้จึงนำ การเรียนรู้ของเครื่อง (Machine learning) ซึ่งเป็นสถิติประยุกต์แบบหนึ่งมาใช้จำแนกหินภายในหลุมเจาะจากข้อมูลดังกล่าว งานวิจัยนี้จะมุ่งเน้นไปที่การทดลองสร้างฟีเจอร์หรือคอลัมน์ใหม่จากข้อมูลเดิมเพื่อพัฒนาความแม่นยำของโมเดลซึ่งยังมีการศึกษาในแง่นี้ไม่มากนักเมื่อเทียบกับการการศึกษาเพื่อหาแบบจำลองที่เหมาะสมที่สุด โดยข้อมูลที่ใช้สร้างแบบจำลองนำมาจาก National Petroleum Reserves in Alaska มีข้อมูล 11 หลุมเจาะ รวมทั้งสิ้นประมาณ 200,000 ข้อมูล จากการสร้างแบบจำลองขั้นต้นโดยใช้แบบจำลองต้นไม้แบบรวมกลุ่ม (Ensemble tree) ซึ่งมีความแม่นยำสูงในหลายการศึกษาในอดีต เพื่อจำแนกหิน 3 ชนิด ได้แก่ หินดินดาน หินทราย และหินปูน ได้ค่า F1-score เฉลี่ย 57.6% และพัฒนาแบบจำลองด้วยการทดลองสร้างฟีเจอร์ใหม่ 4 ฟีเจอร์ ฟีเจอร์แรกคือการสุ่มเพิ่มและสุ่มลดข้อมูลหินบางชนิดเพื่อแก้ปัญหาปริมาณข้อมูลหินแต่ละชนิดที่แตกต่างกัน ฟีเจอร์ที่สองคือการคำนวณค่า M,N จากค่าความหนาแน่น ค่านิวตรอน และค่าหยั่งโดยเสียง ถัดจากนั้น เพื่อให้ค่ารังสีแกมม่าในแต่ละหลุมเปรียบเทียบกันได้ จึงปรับค่ารังสีแกมม่าด้วยวิธีต่าง ๆ 3 วิธีได้แก่ การหาค่ามาตรฐาน (Standardization) การนอร์มอลไลเซชัน (Normalization) การจัดอันดับ (Ranking) และฟีเจอร์สุดท้ายจัดการกับค่าผิดปกติโดยใช้ 2 วิธีได้แก่ การเล็ม (Trimming) และการวินเซอร์ไรส์ (Winsorizing) จากการศึกษาพบว่ายังไม่สามารถสรุปผลกระทบของการสุ่มลดต่อความแม่นยำของแบบจำลองได้ แต่การสุ่มเพิ่มส่งผลให้แบบจำลองมีความแม่นยำลดลง ส่วนค่า M,N เพิ่ม F1-score ได้โดยเฉลี่ย 1% ส่วน การปรับสเกลและการจัดการกับค่าผิดปกติที่เพิ่ม F1-score ได้โดยเฉลี่ยมากสุดประมาณ 1% คือการนอร์มอลไลเซชันและการเล็ม เมื่อนำฟีเจอร์ที่ช่วยพัฒนา F1-score มาใช้ร่วมกันทั้งหมด และทดลองปรับค่าไฮเปอร์พารามิเตอร์ (Hyperparameter) เช่น gamma, max_depth, learning_rate, และ n_estimators จึงสร้างแบบจำลองท้ายสุด ซึ่งได้ค่า F1-score เฉลี่ย 60.5% ในการศึกษาอื่นมีการสร้างฟีเจอร์ เช่น ตำแหน่งของข้อมูลในชุดหิน และหลักฐานบ่งชี้สภาพแวดล้อมการสะสมตัวในทะเล ซึ่งเป็นข้อมูลที่ได้จากการวิเคราะห์โดยนักธรณีวิทยา

ภาควิชา ธรณีวิทยา ลายมือชื่อนิสิต ....

สาขาวิชา ธรณีวิทยา ลายมือชื่อ อ.ที่ปรึกษาหลัก..

ปีการศึกษา 2561 ลายมือชื่อ อ.ที่ปรึกษาร่วม..

5832716123 : MAJOR GEOLOGY

KEYWORDS : WELL LOG INTERPRETATION / MACHINE LEARNING / FEATURE ENGINEERING

THANYABOON SUDHASIRIKUL : DATA MINING OF WELL LOGS USING DECISION TREE BASED MODEL. ADVISOR : ASSIST. PROF. DR.WARUNTORN KANITPANYACHAROEN, Ph.D., 59 pp.

Well-logging is a geophysical survey which provides insights into subsurface geology of an interested borehole. However, the interpretation of well logging data is a time-consuming process and requires an interpreter's experience. Quantitative approaches are attempted to improve time efficiency. This study uses machine learning model which is one of applied statistics to classify well log lithology and focuses on creating new features or columns to improve model performance which is not widely studied while many studies have been focused on choosing the best model. Data are from the National Petroleum Reserves in Alaska and consist of 11 wells which are 200,000 data in total. Ensemble tree model which shows outstanding performacne in previous studies is used to created basic model to classify 3 rock types; mudstone, sandstone, and limestone. The performance of basic model reaches 57.6% of average F1 score and is further improved by incorporating four engineered features. The first feature is known as upsampling and downsampling which is used to manage imbalanced dataset. The second feature involves a calculation of M and N indexes from density, neutron, and sonic logs. To effectively compare and scale data from different wells, the third feature is created through standardization, normalization, and ranking. The fourth feature is developed to reduce data sensitivity and manage outliers by incorporating trimming and winsorizing methods. Results from a combination these features show that upsampling can not improve the model while the effect from downsampling is inconclusive. M and N indexes can slightly improve the model by 1%. The best model combination involves normalization and trimming, which improves the average F1 score by 1%. Hyperparameters of the best model combination such as gamma, max_depth, learning_rate, and n_estimators are tuned to develop the final model, which reaches 60.5% of average F1 score. Further improvement of the classification model can be done by incorporating relative position within a lithologic formation and marine/non-marine indicator.

Department : Geology
Field of Study : Geology
Academic Year : 2018

Student's Signature........ธัญบูรณ์..........
Advisor's Signature........Waruntorn..........
Co-advisor's Signature...................................

Acknowledgements

I would like to express my greatest and sincere appreciation to my advisor Assistant Prof. Dr. Waruntorn Kanitpanyacharoen for her and constructive suggestions all through this research work. Her willingness to give her time so generously has been very much appreciated. My grateful thanks are also extended to our group project; Mr. Pitchaya Hotarapavanon, Ms. Khattiyaporn Tiprongpon, Ms. Ontima Yamchuti, and Mr. Worapop Thongsame for their supporting and useful recommendation.

I would also like to extend my thanks to the Data Cafe Company Limited company where I was interned and gained experiences in data analysis and machine learning.

Finally, I wish to thank all of the Department of Geology's staffs for supporting everything during this study.

Thanyaboon Sudhasirikul
Author

List of Contents

List of Figures

List of Tables

Chapter 1
Introduction and literature reviews

1.1 Background

Geophysical survey and borehole survey are used to explore subsurface geology of an interested area. The first type is a ground-based survey which measures physical properties underneath the Earth surface through the geophysical instruments. For instance, magnetometer detects anomalies, resulting from the presence of ferrous metal-bearing minerals, in a magnetic field at the surface. Geophysical methods such as resistivity survey, airborne magnetic survey, and seismic survey can penetrate through hundreds of meters of depth and collect data over tens of square kilometers. These methods are useful for understanding subsurface structure (Fig 1.1) but limited on detailed subsurface lithology. Another type of survey is a borehole survey. The borehole survey involves well drilling process and uses the drilled wells as the representative of the study area. In contrast to a geophysical survey, each borehole can provide detailed lithology, porosity, and organic content of the area.

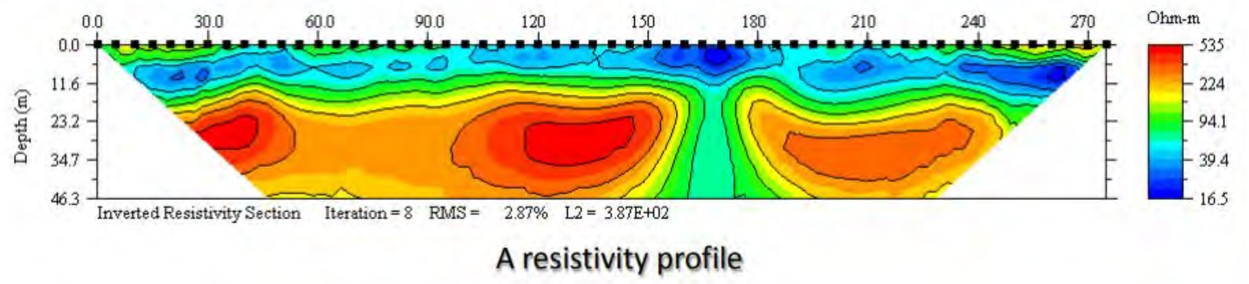


Figure 1.1 Pseudosection of resistivity profile, which is one type of geophysical survey, shows overview and orientation of subsurface geology

(image source: https://www.appstate.edu/~marshallst/GLY3160/lectures/12_Resistivity.pdf)

In borehole survey, coring might be performed in addition to drilling. Coring is a process which cuts the rock within the borehole by using diamond cutting device to obtain a cylindrical core sample. The cores are further studied and analyzed thoroughly to directly identify the subsurface lithology. However, the full-depth recovery of the core is hardly possible due to the expensive cost of drilling/coring and the brittleness of rocks (Fig 1.2). Geophysical information inside a borehole are thus collected by sending specific sensors such as gamma, ???, etc. down the borehole to the depth of interest. This method is known as well logging. Well logging data is interpreted to indirectly identify the subsurface lithology. For instance, an interval of depth which has low gamma-ray is interpreted as sandstone. However, the interpretation of well logging data is a time-consuming process and requires an interpreter’s experience. Well logging has been widely used in natural resource prospecting, particularly for petroleum and geothermal resources, because of its practicality. Coring is generally used as a supporting method, especially in the depth interval that contains poor resolution or conflicting well-logging interpretation.

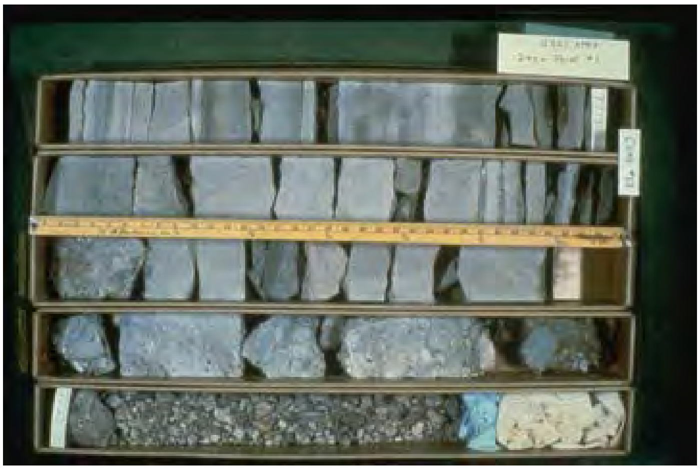


Figure 1.2 Core samples contain missing intervals because the rocks are broken into fragments (USGS, 1999)

Quantitative approaches are attempted to improve time efficiency in well log interpretation (Enikanselu and Ojo, 2012). A study by Busch et al. (1987) applies discriminant analysis on well log data to classify lithology of the Shublik formation in North Slope, Alaska, USA,. The analysis used M and N indexes which are calculated from density, neutron, and sonic logs as independent variables to classify 3 rock types; limestone, shale, and sideritic mudrock. Results from discriminant analysis report 70-80% accuracy of lithology classification. Even though the conditions are limited to only 1 formation at a time, the analysis shows a lot of promise. A study by Dubois et al. (2007) uses well log data from Panoma field in Southwest Kansas, USA to compare conventional statistical analyses with machine learning models. Statistical analysis often assumes a specific distribution of input data meanwhile machine learning, which is a form of applied statistics, concerns less about the distribution. Due to the high dimensionality and nonlinear-relationship of the data, the assumptions do not perfectly hold for the statistical analysis and contribute to inferior accuracy or performance as the study demonstrates.

Each datum for a classification machine learning model consists of 2 parts. The first part is a label, which describes a class or group of each datum. The second part is features or independent variables, which describe a set of measured properties. The model assumes that each label has a unique pattern of variables and finds underlying patterns to classify the label. Error or performance of the model is evaluated, then the machine tries to revise the model repeatedly to minimize the error; for instance, adjusting the variables' weight in logistic regression or recalibrating the cluster centroid in k-means clustering. Using different machine learning models; for example, k-nearest neighbor, support vector machine, random forest, and artificial neural network can lead to a significant difference in model performance (Dubois et al., 2007, Hall and Hall, 2017). Furthermore, preprocessing of variables also contributes to the improvement of model performance (Bestagini, 2017; Chen and Zeng, 2018). The preprocessing is also known as feature engineering. For example, a logarithm of resistivity log divided by the logarithm of neutron-density or M and N index from M-N crossplot are examples of the engineered feature.

Decision tree is a machine learning model which uses a flowchart-like structure to classify the data (Navada et al., 2011). Its structure resembles human decision making. However, it tends to create a complex flowchart and overfit with the training data, so the ensemble tree model is invented (Breiman, 2001; Chen and Guestrin, 2016). The ensemble tree model is made of many simple decision trees. Those simple trees are constructed differently, so they might classify the data differently. Majority voting is used to conclude the classification process. A study by Hall and Hall (2017) shows that the ensemble tree has the highest classification accuracy among other models such as support vector machine, deep neural network, k-nearest neighbors, multilayer perceptron, convolutional neural network.

Through considerable development on both the model and feature engineering, machine learning could perform some task at the same level as human or even higher; for example, legal contract reviews, clinical image diagnosis, and playing Go (Silver et al., 2017; Loh, 2018; LawGeex, 2018). Remote sensing also adopts a machine learning model in image classification (Prasad et al., 2017). In summary, machine learning is a reliable tool to assist human in many fields.

In this study, well log data are collected from the U.S. Geological Survey database of National Petroleum Reserves in Alaska. This project thus aims to develop a classification model based on the ensemble tree models. Data are preprocessed or feature engineered in various methods to study their effects on the model performance.

1.2 Objectives

1.2.1 To develop ensemble tree model for well-logging lithology classification

1.2.2 To measure the effect of engineered features on the model performance

1.3 Benefits

1.3.1. Own Experiences

- Learn basic well logging interpretation
- Learn statistical analysis and machine learning development including exploratory data analysis, feature engineering, machine learning model development, and model evaluation
- Learn scientific research methodology and presentation

1.3.2 . Society

- Develop effective feature engineering for well-logging lithology classification

1.4 Literature reviews

1.4.1 Machine learning

Machine learning is the use of statistics and computer algorithm to learn from data and create model to complete some task without being explicitly programmed (Barber, 2007). In other word, the problem is given to the machine, and it figures out detailed solution on its own by using the designed method. Because machine learning is an application of statistics, they both uses the same form of data, which is table-based. There are two main types of machine learning; supervised and unsupervised. For the supervised, each row or datum consists of two parts; label and feature(s). The feature part is a set of independent variable which describes measured or observed properties while the label is an dependent variable which expresses the outcome of those feature. The numerical label leads to regression model and the categorical label leads to classification model (Fig 1.3). In contrast, the unsupervised machine learning does not have the label, so it focuses on the discovery of clusters or links from the features alone (Fig 1.4).

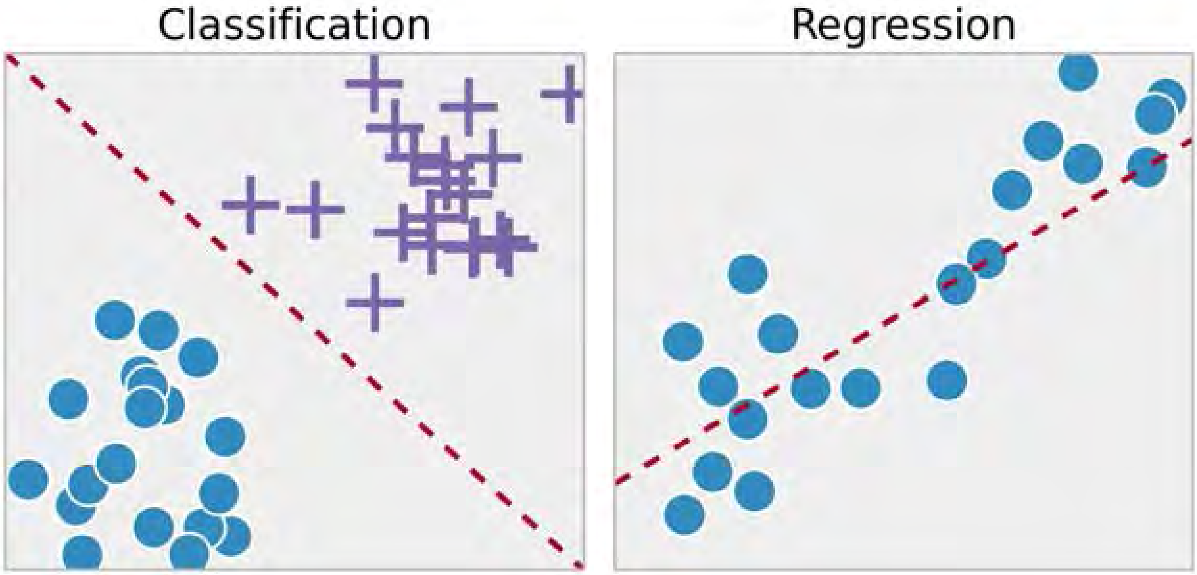


Figure 1.3 Examples of the classification model and the regression model

(image source: https://cdn-images-1.medium.com/max/1600/0*WE3Sz--1NUEWBmUR)

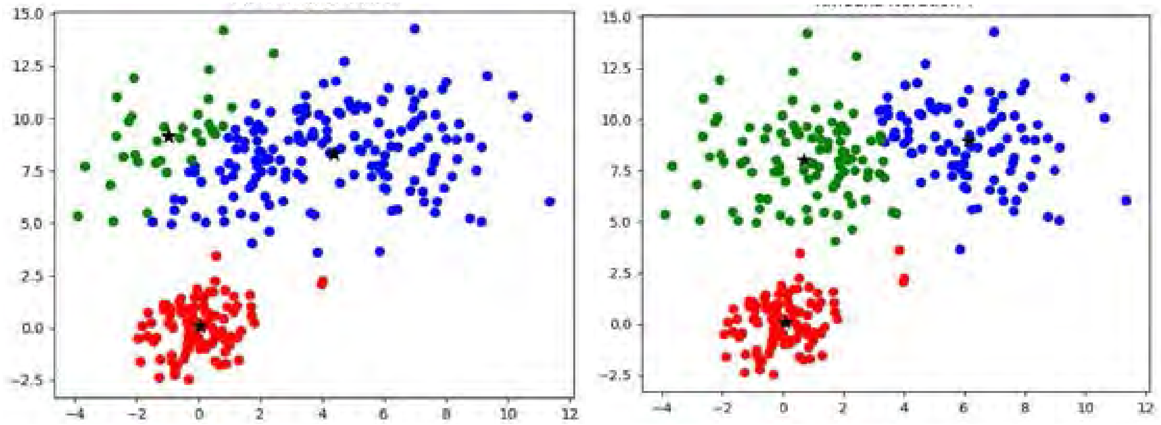


Figure 1.4 An example of clustering model shows re-clustering process to locate the most appropriated centroid point

(image source: https://sandipanweb.files.wordpress.com/2017/03/kmeans8.gif?w=676)

After the model has learned, its accuracy will be tested and evaluated by the test set which has not been input to the model before. There are many evaluation metrics for classification problem, but the metrics from studies related to this study will be described here. First, accuracy is the most basic evaluation metric. It describes percentage of total correctly classified data. Accuracy is simple to understand, but on unbalanced data, it does not reflect the performance of the model very well. An example of unbalanced data is a dataset which has 5% of data labelled as “positive” and 95% of data labelled as “negative”. The model can show accuracy as high as 95% when it classify all test data as “negative”. To correctly evaluated the model performance in that situation, precision and recall are introduced (Kent et al., 1955). The metrics evaluate the model performance on each label separately. In binary classification, positive and negative, precision is defined as;

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \qquad \text{... (Equation 1)}$$

where True positive is the amount of positive which is classified as positive, False positive is the amount of negative which is classified as positive. In other word, precision is a percentage of correct classification from the data classified as one label. Meanwhile, recall is defined as;

$$Recall = \frac{True\ positive}{True\ positive + False\ Negative} \qquad \text{... (Equation 2)}$$

where False negative is the amount of positive which is classified as negative. In other word, recall is a completeness in finding one label.

1.4.2 Decision tree model

Decision tree is a machine learning model which uses a flowchart-like structure to classify the data (Navada et al., 2011). This model is first introduced in a study by Morgan and Sonquist (1963) and is a foundation of more complex models such as Random Forest (Breiman, 2001) and XgBoost (Chen and Guestrin, 2016). In order to develop the decision tree model, training data are input into the model. The label describes a class or group of each datum, and the model assumes that each label has a unique pattern of variables and finds underlying patterns to develop a flowchart for classifying the label.

The flowchart has 2 types of node; splitting node and leaf node (Fig 1.5). The splitting node is a condition to check on the data which is based on the underlying patterns and is used to decide which path a datum should flow through. Any path leads to a conclusion or label is called a leaf node, which classifies the label for the datum.

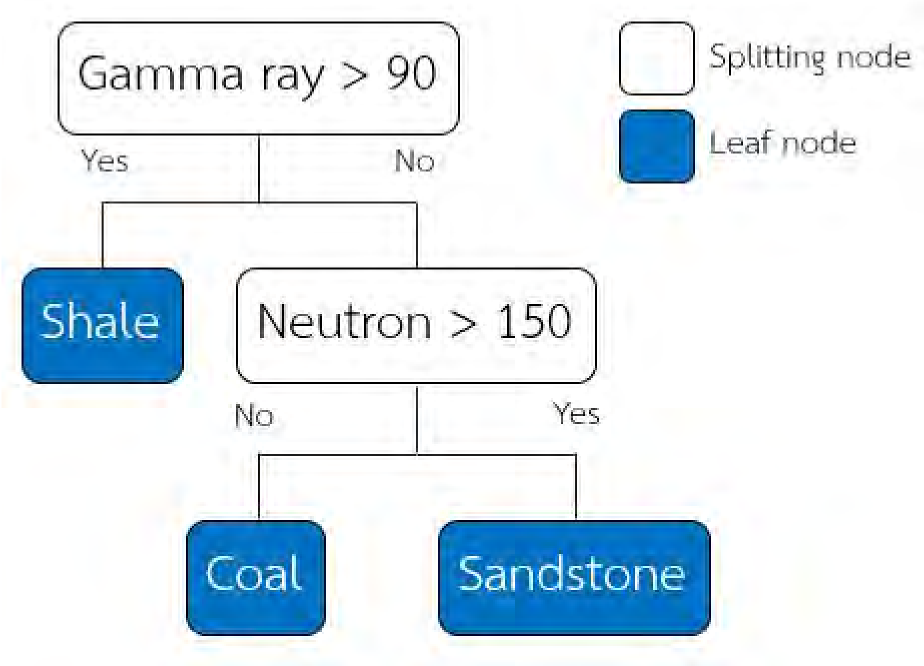


Figure 1.5 Example of the decision tree model. If gamma ray of the datum is more than 90, the model classifies it as Shale. If the datum doesn't meet the condition, it will be considered further by Neutron.

The flowchart might have many splitting nodes which related to each other hierarchically. To represent the underlying pattern, each splitting node must has the right feature, including its value, at the right level of the hierarchy. Homogeneity of training data before splitting is measured, then the decision tree model defines "right" as the highest improvement in homogeneity after splitting into 2 groups. Homogeneity could be measured by several metrics such as Gini index and entropy. The entropy is defined as (Shanonn, 1948);

$$H = -\sum_{i=1}^{N} P(x_i) \cdot log_N P(x_i) \qquad \text{... (Equation 3)}$$

where H is Entropy of the data, N is the number of labels within the data, and $P(x_i)$ is a proportion of the number of datum with label x_i to the number of total data. Entropy measures impurity of the data, so single-label data has 0 entropy while the entropy of N-labels data with an equal number of each label is 1. Reduction in the entropy is highly correlated to improvement in the homogeneity.

Likewise, the Gini index is defined as (Ceriani and Paolo, 2011);

$$G = 1 - \sum_{i=1}^{N} P(x_i)^2 \qquad \text{... (Equation 4)}$$

where G is Gini index of the data, N is the number of labels within the data, and $P(x_i)$ is a proportion of the number of datum with label x_i to the number of total data. However, the Gini index is on a scale of 0 to 1-1/N which is ranging from the highest to lowest homogeneity.

A study by Kotsiantis (2011) shows that decision tree model is widely accepted and used in practice because it is not restricted by the distribution of training data and uses reasonable time to classify a large amount of data. Also, the model resembles human decision making, so it is simple to understand and interpret. However, it has a tendency to overfit with the training data by constructing a complex or long flowchart. Thus, the model might classify the training data with almost 100% accuracy, but the accuracy drops significantly with the new data or test data (Fig 1.6). This problem is commonly known as overfitting.

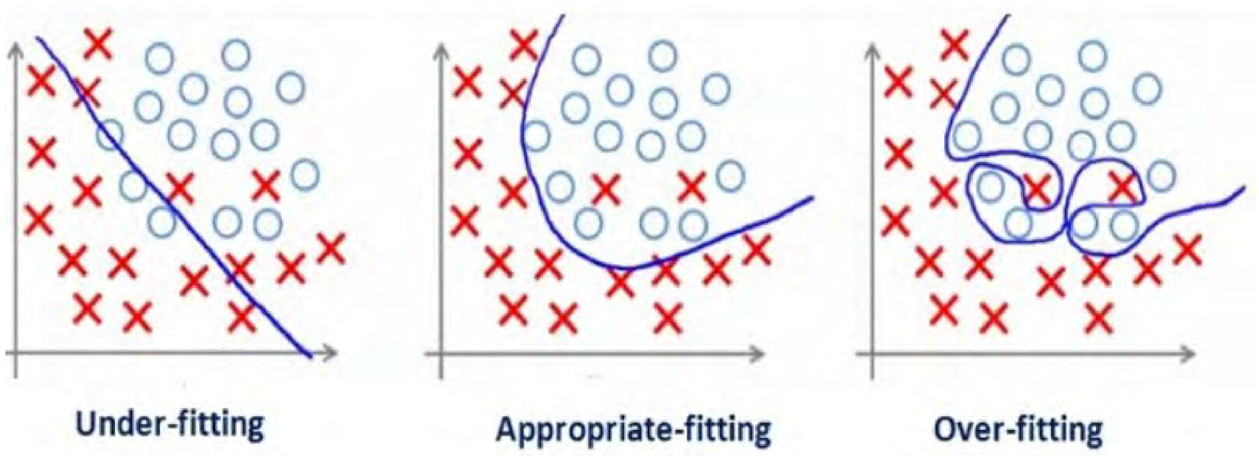


Figure 1.6 Different types of result in machine learning; underfitting, desired, overfitting. Underfitting model is too simple to explain the data, while overfitting model is too accurate and unrealistic.

(image source: https://cdn-images-1.medium.com/max/1600/0*7xAFG32QA2nNEs6n)

To overcome the overfitting problem, the length of the tree is limited, but multiple trees are constructed and made a decision together (Fig 1.7). The same level of accuracy in training data is obtained together with better accuracy in the test data. The technique is called ensemble method. Examples of ensemble tree model are Random Forest (Breiman, 2001) and XGBoost (Chen and Guestrin, 2016). Random Forest uses a bagging ensemble method, which randomly repeats sampling the training data for each tree. In contrast, XGBoost uses a boosting ensemble method, which creates a new tree based on misclassification of the previous tree.

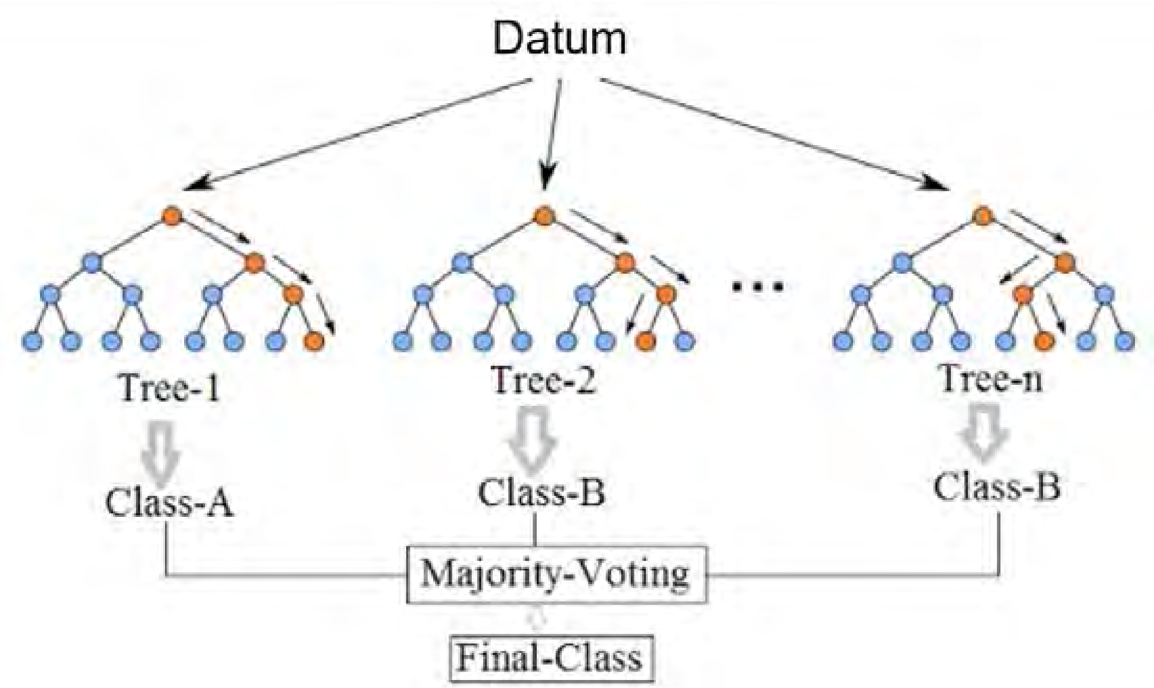


Figure 1.7 Ensemble tree model is composed of multiple decision trees. Each tree is different, so they might not classify a datum as the same label.
(image source: https://www.kdnuggets.com/wp-content/uploads/rand-forest-1.jpg)

1.4.3 Hyperparameters Tuning

The machine learning model is being widely used, so many pre-built modules are developed; for instances, scikit-learn python library (Varoquaux et al., 2011). They provide standardized models which can be used instantly. However, the setting of the model could be configured as well; for examples, a number of trees and homogeneity metric in an ensemble tree model. The setting is called hyperparameters in distinction to parameters which are used to describe weight or coefficient of each independent variable in a mathematical function. In fact, the hyperparameters specify details of how the model learn and solve a problem, and they are contributed to the accuracy of a model too. Examples of hyperparameters are a number of estimators and minimum impurity decrease. A number of estimators specify the number of trees in the ensemble model while minimum impurity decrease sets a threshold for the decision tree to split the node

1.4.4 Well log lithology classification and statistical approach

A study by Busch et al. (1987) examines well logging lithology classification using discriminant analysis, which is a type of statistical analysis. The study focuses on Shublik Formation of the Prudhoe Bay, North Slope, Alaska. Neutron, sonic, density, and gamma-ray logs are available, so M and N indexes are calculated. M index is a combination of sonic and density logs while N index is a combination of neutron and density logs. These M and N indexes are good lithology indicators. First, only M index is considered. A total of 3 rock classes; sideritic mudrock, shale, and limestone show distinctive distributions (Fig 1.8, left), so the classification boundary could be constructed at the intersection of each pair of the classes (Fig 1.8, right). The analysis obtains 76.20% accuracy.

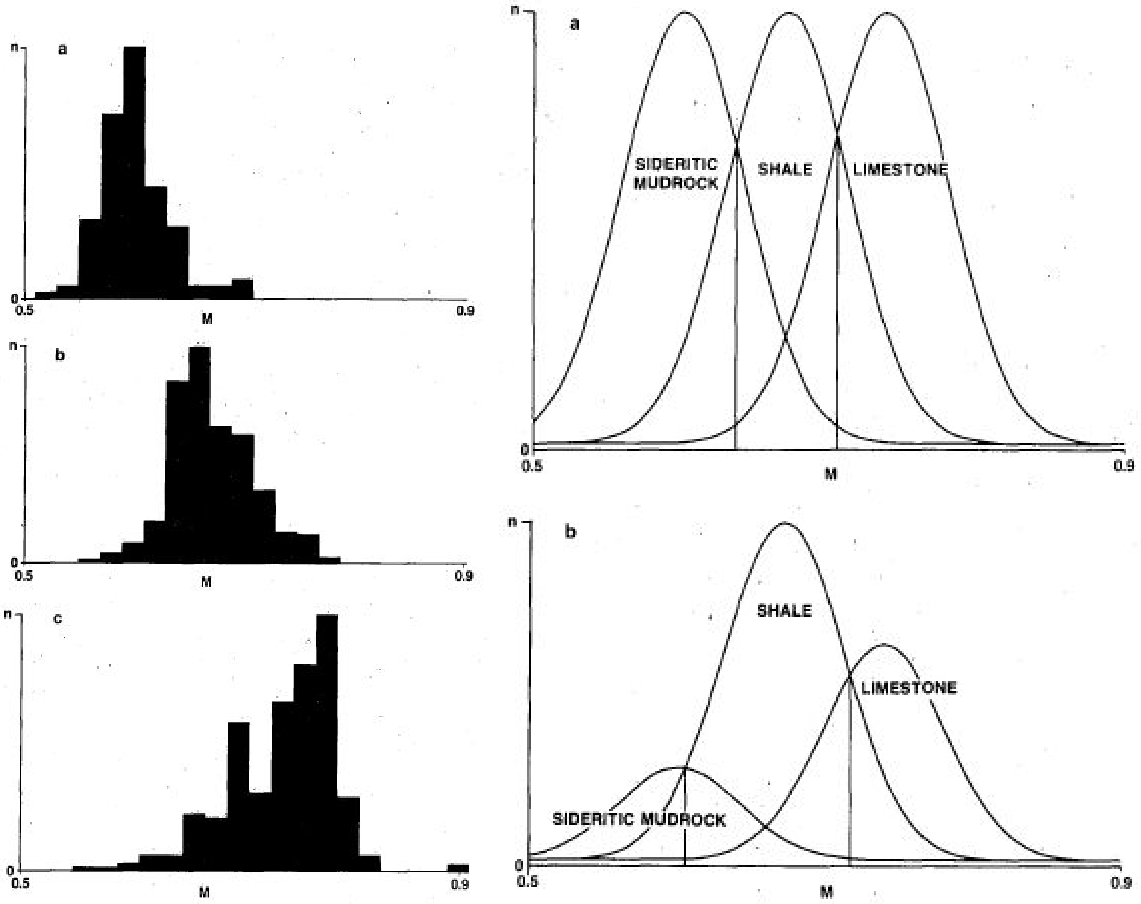


Figure 1.8 (Left) Histograms of M (Busch et al., 1987); (a) sideritic mudrock (b) shale (c) limestone (Right) Distribution function of M (Busch et al., 1987); (a) regular discriminant analysis (b) discriminant analysis which is weighted by occurring proportion. Classification boundaries are constructed vertically between sideritic mudrock-shale and shale-limestone.

Then, the full analysis is conducted which includes all 7 rock classes and uses both M and N index. The full analysis also uses “relative position”, which describes a position of datum relative to the bottom of Shublik Formation. The model is able to obtain 77.57% accuracy in the validation data and 75% in the test data. Even though the method is designed to handle a single rock formation at a time, the test accuracy which is as high as the validation accuracy ensures robustness of the method. In addition, almost all of the misclassification could be explained by thin-bed effects, log resolution problems, and core-log misalignment problems.

A study by Dubois et al. (2007) compares 4 classification models from both conventional statistical analysis and machine learning. These models are discriminant analysis with Bayes’ rule, fuzzy logic, k-nearest neighbor, and artificial neural network. Geologic observations from core samples, which are marine/nonmarine and relative position, are also included in addition to well log measurements; gamma-ray log (GR), resistivity log (ILD_log10), photoelectric log (PE), average neutron-density log (PHIND), and neutron-density difference (DeltaPHI). Artificial neural network shows the best accuracy performance (Fig 1.9). The discriminant analysis does not perform well because the data has high dimensionalities and non-linear relationships. In addition, the data are often overlapped, which lead to poor performance of this simple model. Even though fuzzy logic and k-nearest neighbor perform well, they are still behind the neural network.

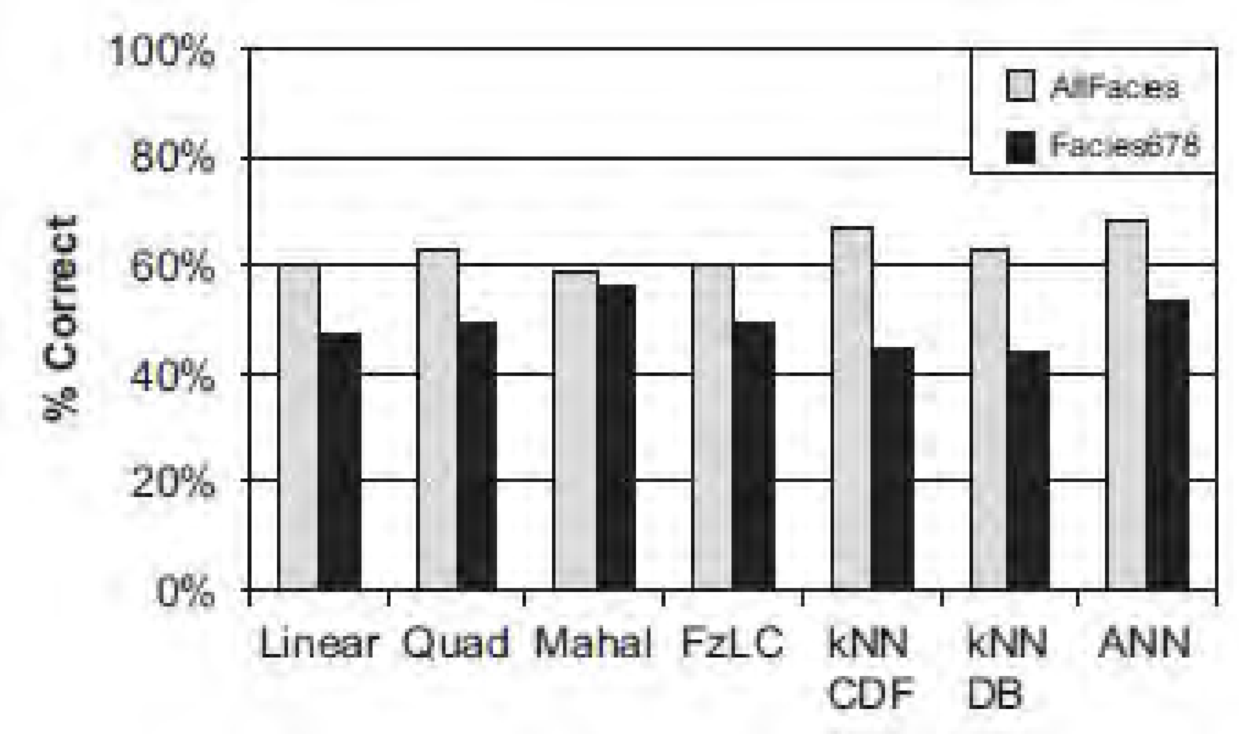


Figure 1.9 Result of 4 classification models (Dubois et al., 2007); discriminant analysis (linear, quadratic, Mahalanobis), fuzzy logic, k-nearest neighbor (cumulative density function and degree of belonging), and artificial neural network.

A study by Hall (2016) demonstrates the performance of a memory-intensive machine learning model called support vector machine (SVM) in well logging lithology classification problem. The same dataset from a study by Dubois et al. (2007) is used. The SVM tries to construct boundary lines to separate the data according to the label. It might even project the data from the original feature or dimension into a higher dimension to aid classification which causes it to uses a lot of computational power. Finally, 43% F1 score is obtained (Fig 1.10). F1 score is an evaluation metric for classification problem which takes Type I and Type II error into account. In other words, it is more considerate than accuracy. The score is then set as a baseline for 2016 machine learning contest which is held by the Society of Exploration Geophysics (SEG).

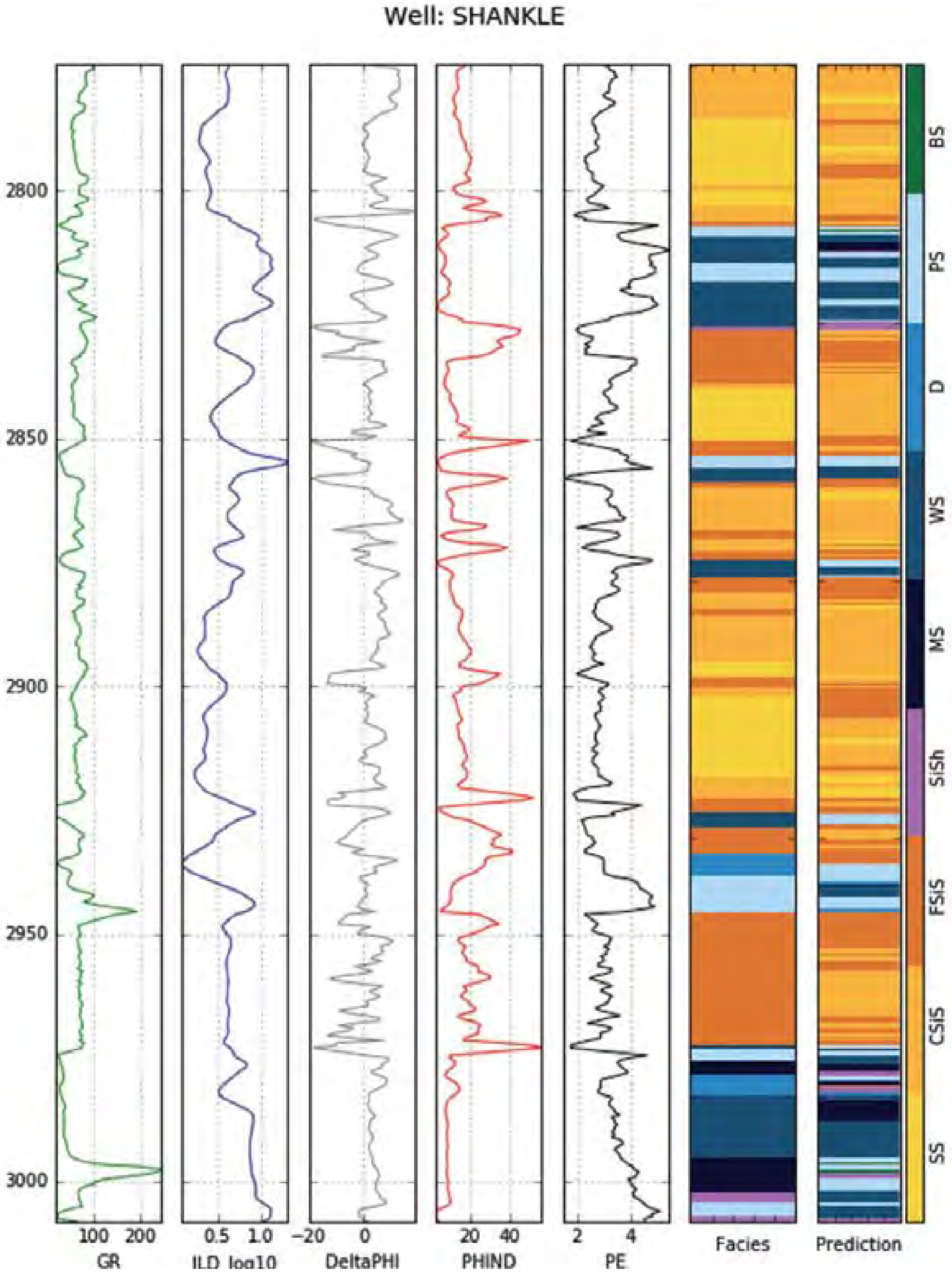


Figure 1.10 The example of well logging lithology classification by the SVM (Hall, 2016). A total of 9 labels are phylloid-algal bafflestone (BS), packstone-grainstone (PS), dolomite (D), wackestone (WS), mudstone (MS), marine siltstone and shale (SiSh), nonmarine fine siltstone (FSiS), nonmarine coarse siltstone (CSiS), and nonmarine sandstone (SS)

A study by Hall and Hall (2017) reports the result of SEG machine learning contest on well log data in 2016. Among 40 participating teams, the winner team, LA team, uses an ensemble tree model which shows and obtains 64% F1 score (Fig 1.11). They also preprocess or feature engineer the data by calculating the gradient of well log measurements at each depth. The idea of the engineered feature comes from the fact that rocks are deposited in form of bedding or interval, so, the relationship of adjacent data should also be considered. In fact, almost all of the top half of the participating team use an ensemble method. The result clearly points out superior performance of an ensemble method over other models such as a neural network, k-nearest neighbors, support vector machine, and majority voting.

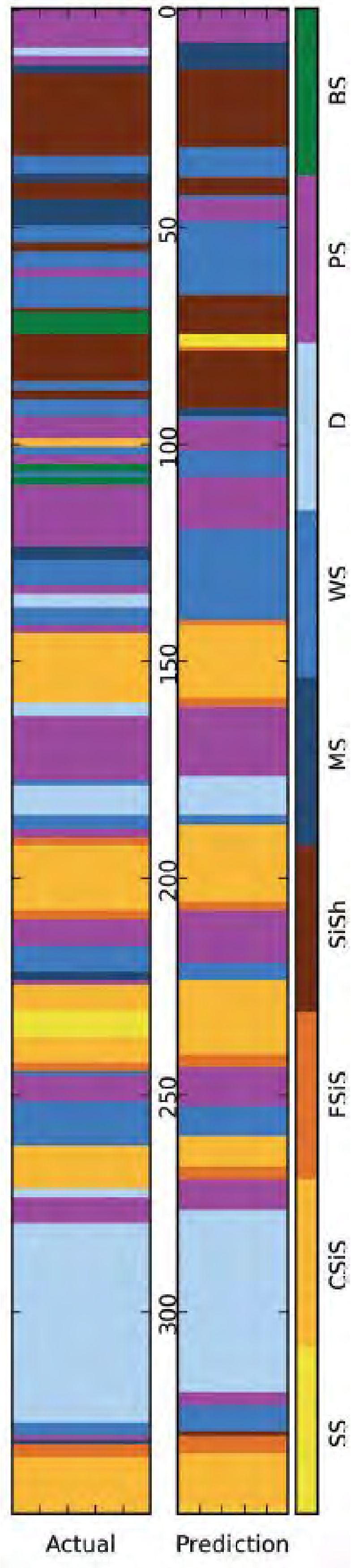


Figure 1.11 The result of LA team in SEG machine learning contest 2016 (Hall and Hall, 2017)

A study by Bestagini et al. (2017) examines the solution of SEG 2016 contest winner further. The study uses the same model but 2 sets of engineered features are generated in addition to the gradient of well log measurements at each depth. The first set is square of each well log measurement. The other set is pairwise multiplications between any 2 well log measurements; for examples, GRxPE and ILD_log10xPHIND. Many engineered features have no physical meaning in Geology and they are even worse when they are not very informative for classifying rocks. However, some of them might be useful, and the ensemble tree model could avoid selecting those meaningless and non-informative by itself during its training. Therefore, the study is a trial-and-error on these arithmetic features and obtains 61% F1 score. Without any feature engineering process, the ensemble tree model alone could obtain at best 55% F1 score.

A study by Chen and Zeng (2018) also revises the solution of SEG 2016 contest winner. However, the engineered feature in the study derives from Archies' equation (Archie, 1942) which is widely used in well log interpretation by a geologist. The equation describes a relationship between resistivity and porosity, it is defined as;

$$F = R_0/R_t = C \cdot p^{-m}$$

where F is called formation resistivity, R_t is the resistivity of the rock saturated with oil and water, R_0 is the resistivity of fluid inside the rock, C is the tortuosity constant, m is the cementation factor, and p is porosity of the rock. The formation resistivity is measured in a resistivity log and the porosity is directly reflected from a neutron-density log. This equation is transformed by taking logarithm function into;

$$log_{10}F = log_{10}C - mlog_{10}p$$

It has a similar form to linear equation $y = mx + b$ where $log_{10}F$ is y-axis variable, $log_{10}C$ is y-intercept value, m is a slope, and $log_{10}p$ is x-axis variable. Since, the study assumes that the cementation factor is unique for each label (Fig 1.12) and the tortuosity constant of each datum could not be measured by well-logging, the cementation factor is then approximately calculated from the proportion of logarithm of resistivity log to logarithm of neutron-density log. The approximated cementation factor contributes to 5% improvement in the F1 score.

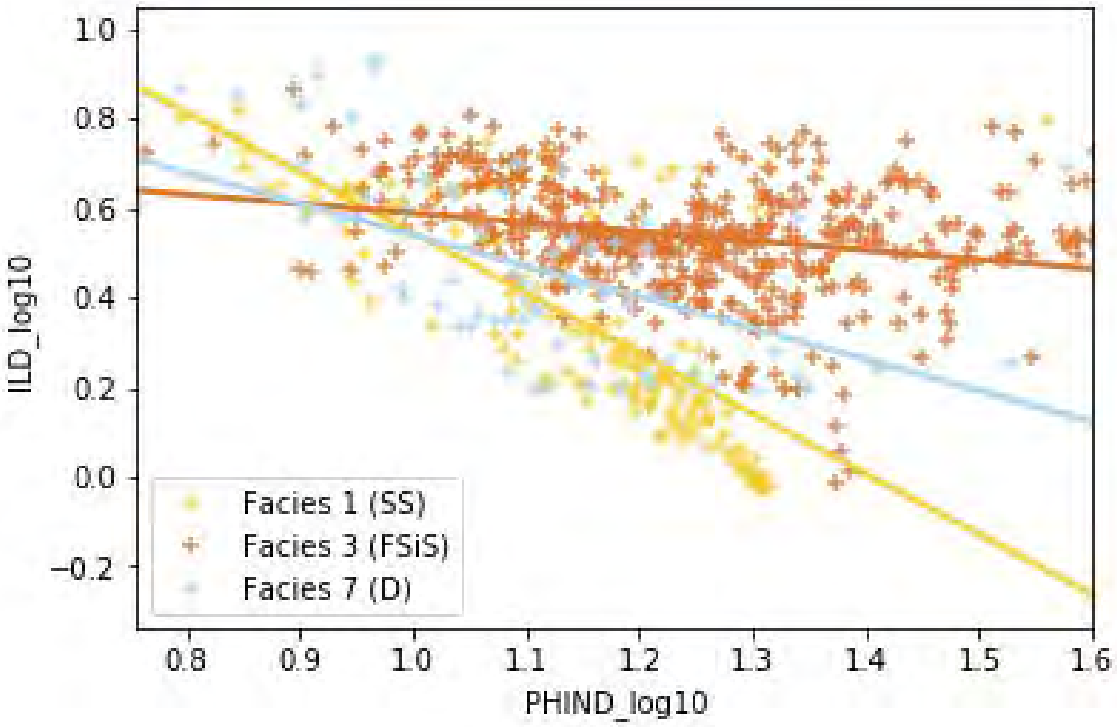


Figure 1.12 Scatter plot of the logarithm of neutron density log and the logarithm of resistivity log (Chen and Zeng, 2018) shows linear trends in nonmarine sandstone (SS), nonmarine fine siltstone (FSiS), and dolomite (D).

Chapter 2
Study Area

2.1 National Petroleum Reserves in Alaska

National Petroleum Reserves in Alaska (NPRA) is continental land on the Alaska North Slope in Northern Alaska, USA (Fig 2.1). It is owned by the United States federal government and managed by the Department of the Interior, Bureau of Land Management. It has an area of 100,167.79 square kilometers. The NPRA is first established and explored in 1923 with a purpose for navy activities. The NPRA is estimated to contain 896 million barrels of conventional, undiscovered oil and 53 trillion cubic feet (1,500 cubic kilometers) of undiscovered, conventional, natural gas (USGS, 2010).

Figure 2.1 The location of Alaska, USA
(Image source: https://commons.wikimedia.org/wiki/File:Map_of_USA_AK_full.svg)

Studies by Wahrhaftig (1965) and Moore et al. (1994) divide the topography of Northern Alaska into 3 provinces; Arctic Mountains, Arctic Foothills and Arctic Coastal Plain from south to north (Fig 2.2). The Arctic Mountains province has linear mountain ranges, ridges and hills at high altitude on the east side, about 3,000 meters above sea level, then their altitude decline westward. These mountain ranges are also called the Brook Range orogenic belt (Fig 2.3). Next, the Arctic Foothills consists of rolling hills, mesas, east-west ridge. The altitude is between 250 - 550 meters which gradually decrease northward. Last, The Arctic Coastal Plain spreads out succeeding the northward trend. The North Slope is composed of the Arctic Foothills and the Arctic Coastal Plain. Northern Alaska is composed of 2 litho-tectonic terranes; Arctic Alaska and Angayucham (Silbering et al., 1994). The term "terrane" is defined as an area bounded by a fault which has unique stratigraphic sequence and geologic history (Jones, 1983; Howell et al., 1985). The North Slope and parts of the Arctic Mountains province are in Arctic Alaska terrane.

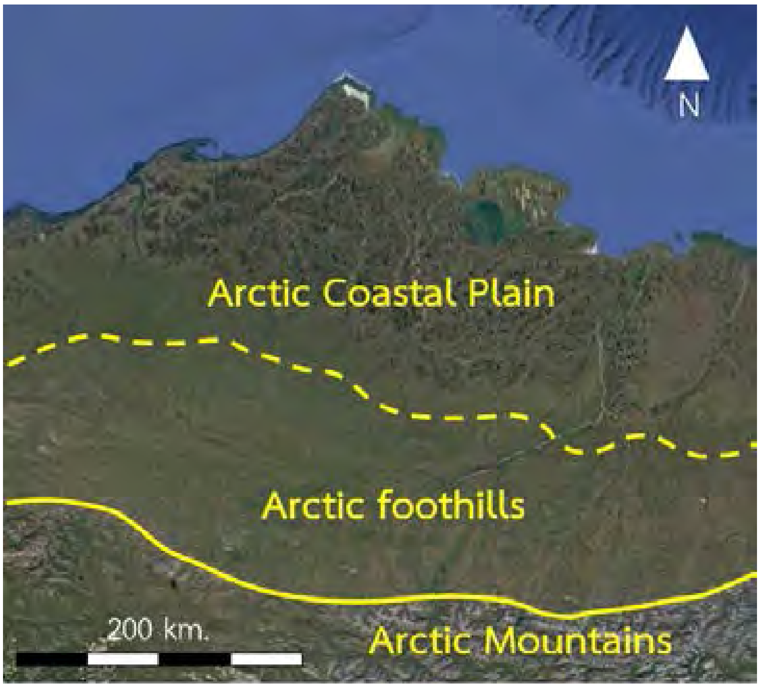


Figure 2.2 The NPRA Map showing 3 topographic provinces; Arctic Mountains, Arctic Foothills and Arctic Coastal Plain

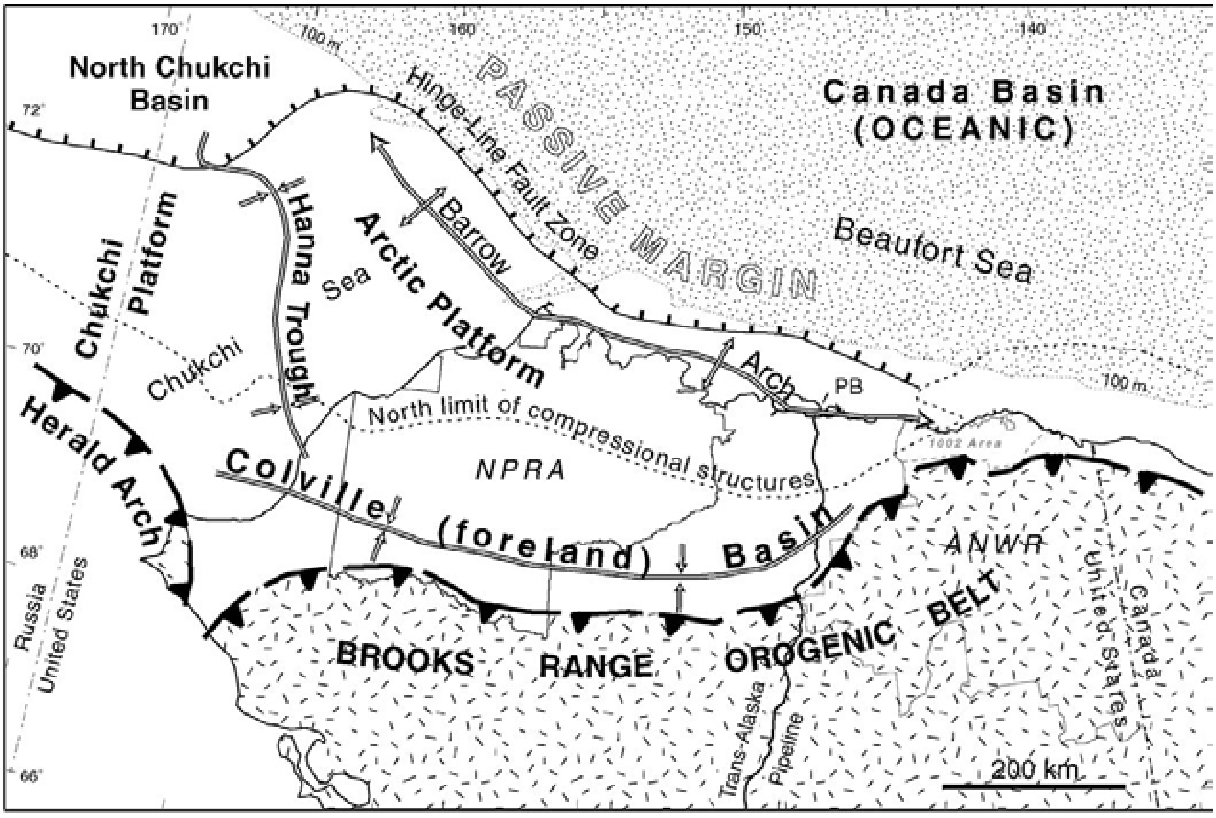


Figure 2.3 The NPRA map showing related showing major tectonic features (Bird, 2001)

2.2 Stratigraphy

A study by Lerand (1973) suggest to group rock of the NPRA into tectonostratigraphic sequences, later, it is revised in accordance with new evidence and studies (Bird, 2001). Therefore, the NPRA is composed of 4 tectonostratigraphic sequences; Frankalinian Sequence, Ellesmerian Sequence, Beaufortian Sequence, and Brookian Sequence (Fig 2.3). First of all, the Frankalinian Sequence is the oldest and considered as a basement rock. It has a variety of rock origin and complex geologic history; ranging from steeply dipping fine-grained marine sedimentary rocks, gently dipping nonmarine sedimentary rocks, igneous rocks, and metamorphic rocks. The sequence’s thickness is increased southward from 4,000 feet in the coast area to 30,000 feet in the northern part of the Arctic Mountains province. It is separated from the Ellesmerian Sequence with regional angular unconformity.

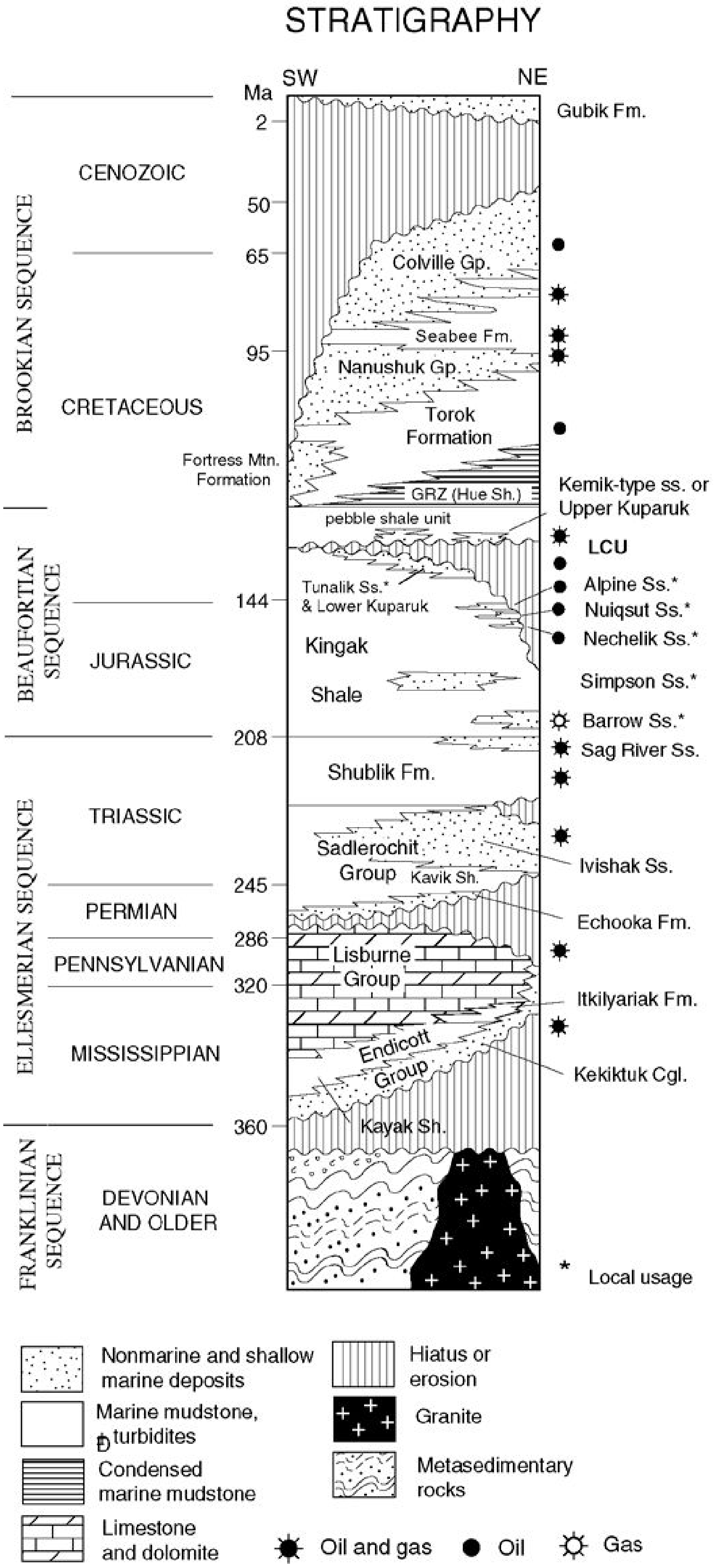


Figure 2.4 Stratigraphic column of the NPRA with 4 tectonostratigraphic sequences and lithostratigraphic units (Bird, 2001)

Second, the Ellesmerian Sequence has carbonate and clastic continental shelf deposit, so its depositional environment is a passive margin. The sequence's thickness is mostly less than 6,000 feet. Seismic reflection shows 3 transgressive-regressive cycles (Bird and Molenaar, 1987); Mississippian to Early Permian, Early Permian to Early Triassic, and Middle Triassic to Late Triassic. The first cycle has nonmarine and shallow-marine clastic rocks, known as Endicott Group, and marine carbonate rocks, known as Lisburne Group. Sedlerochit Group represents the second cycle; marine sandstone of Echooka Formation, prodelta marine mudstone of Kavik Shale, and deltaic sandstone of Ivishak Formation. The last cycle has organic-rich calcareous mudstone and marine sandstone which are named Shublik Formation and Sag River Sandstone, respectively. Shublik Formation is the major source rock for petroleum system in the NPRA.

Next, the Beaufortian Sequence is a stratigraphically complex which is classified as a syn-rift deposit during Jurassic to Cretaceous (Hubbard et al., 1987). The sediment supply is from the north or local pre-rift rocks. The lithology is a marine mud-dominated strata with local sandstone which are named Kingak Shale. Last, the Brook Range orogenic uplift supply Cretaceous to Tertiary sediments for the Brookian Sequence. It can reach maximum thickness over 25,000 feet in an area near the source; for instances, Colville foreland basin. The deposit overwrites the old rift shoulder of Beaufortian Sequence and forms parts of a passive margin, continental shelf and slope, for the North Slope. Fortress Mountain Formation is the most proximal unit which the outcrops show at the Arctic Foothills province. Various units are founded along the northward path of sediments in the passive margin; deltaic deposits of Nanushuk and Colville Group, shelf mudstones and turbiditic sandstones of Torok Formation and Seabee Formation, condensed marine mudstone of Hue Shale. Hue Shale is a common source rock in a shallow zone of the NPRA which is remarked by the high gamma-ray log, the gamma-ray zone (GRZ).

2.3 Tectonic setting

Tectonic setting of the North Slope is summarized as follows; the North Slope is bounded on the north by the passive continental margin of the Canada Basin and on the south by the Brooks Range orogenic belt. The boundary between Late Devonian of Frankalinian Sequence and Early Mississippian shows compressional deformation along the shoreline. Granitoid plutonism also presents during the same period. Next, sediments are deposited in passive margin environment until Late Jurassic before rifting of the Arctic Ocean basin begins. The Canada basin margin shows a record of an extensional structure where the rifting occurred til Early Cretaceous. Since Early to Middle Eocene, the North Slope rotates in counterclockwise. As a result, thrusting and uplifting occur on the southern part and supply sediments to the Colville foreland basin.

Chapter 3
Methodology

The study of engineered features on the ensemble tree model for well-logging lithology classification progresses through many steps; data collection, exploratory data analysis, model development, respectively (Fig 3.1).

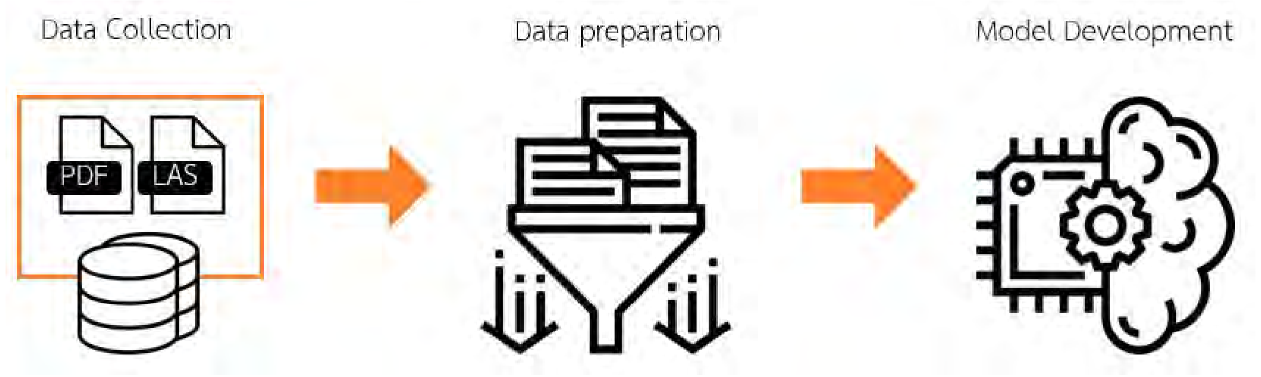


Figure 3.1 Three main steps of this study

3.1 Data collection

There are a total of 218,038 data points collected from 11 wells. The depth of well ranges from 600 km. to 6,000 km. The types of well-logging include gamma-ray, deep-induction resistivity, medium-induction resistivity, shallow-induction resistivity, neutron, density, sonic, self-potential, caliper, laterolog deep, laterolog shallow. Each well has different types of well-logs but contains 8 logging in common. Well-logging has operated only the depths where it is economics and necessary because it requires resources such as money and time. Hence, some wells contain missing values, and they are removed from the analysis. Missing values account for 40% of the total amount of data. The final amount of data 127,784 well-logging data points are used in this study.

The data contain 6 rock types; mudrock, sandstone, siltstone, limestone, coal, and igneous rocks (Fig 3.2). Mudrock is the most common, and the second most is sandstone. Both of them appear in every well. Coal is the most uncommon, it is so rare that the amount is less than 1% of the mudrock. The rest can only be found in some wells.

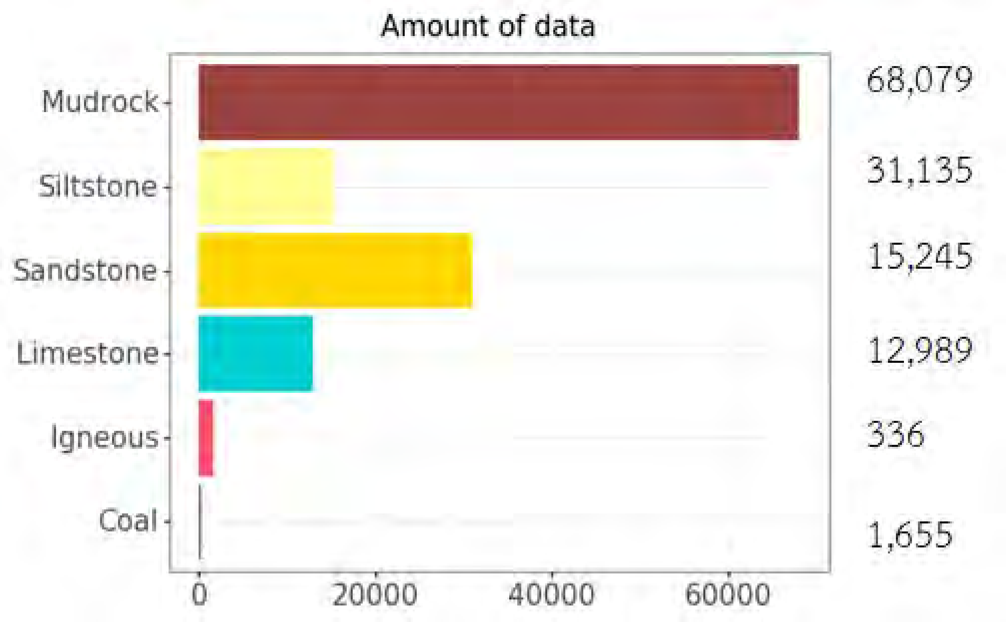
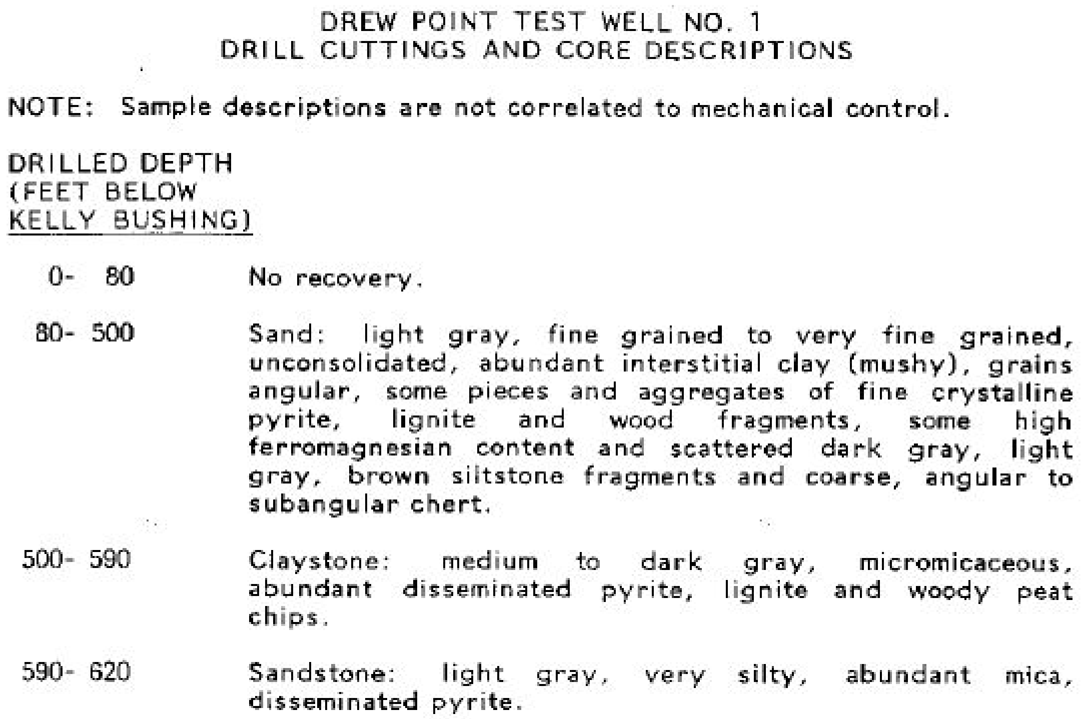


Figure 3.2 Rock types which are used in this study are plotted in a barplot. The Mudrock rock type has more than 60,000 data while the Coal and Igneous rock type have less than 2,000 data, so the dataset is quite unbalanced.

In order for the model to learn the pattern of rock types and automated classify, any interval of well log measurements must be labeled with the rock type beforehand. Hence, core descriptions are the most suitable and accurate source of information because they are derived directly from rocks inside the well. Core descriptions are provided in Portable Document Format (PDF) file which is scanned from the printed report (Fig 3.3), so they are images of texts, not the text itself. The information from core descriptions is thus challenging to extract.



DREW POINT TEST WELL NO. 1
DRILL CUTTINGS AND CORE DESCRIPTIONS

NOTE: Sample descriptions are not correlated to mechanical control.

DRILLED DEPTH
(FEET BELOW
KELLY BUSHING)

0- 80 No recovery.

80- 500 Sand: light gray, fine grained to very fine grained, unconsolidated, abundant interstitial clay (mushy), grains angular, some pieces and aggregates of fine crystalline pyrite, lignite and wood fragments, some high ferromagnesian content and scattered dark gray, light gray, brown siltstone fragments and coarse, angular to subangular chert.

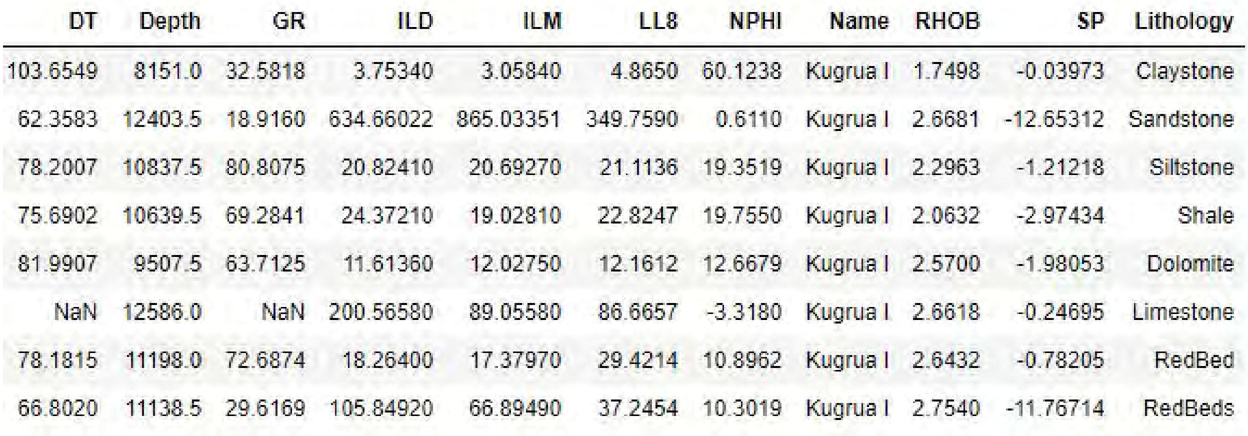
500- 590 Claystone: medium to dark gray, micromicaceous, abundant disseminated pyrite, lignite and woody peat chips.

590- 620 Sandstone: light gray, very silty, abundant mica, disseminated pyrite.

Figure 3.3 An example of core description from Drew Point I well

A Python module named Lasio is used to access the well log measurements and used them as features in the machine learning. Meanwhile, to collect the label conveniently, many steps are needed. First of all, the text data are extracted from the core description by optical character recognition (OCR) from Google Cloud vision application programming interface (API). The text data composed of an interval of depth followed by a description of rocks in that interval. The description also has repetitive form; a rock type then its physical observation. In almost all of the descriptions, the major rock type is described first and the minor are followed. In this study, the major rock types which are mentioned first are treated as a representative of each interval. With the recognizable pattern, it is possible to extract the interval of depth together with its major rock type automatically by Regular Expression (REGEX) which is one of a programming language.

The REGEX utilizes the fact that text data might have a recognizable form so a defined sequence of character could capture them. Some of the label or rock type are mistyped and letter-case sensitive so they are taken care of afterward. Finally, the labels consisted of these following; Mudrock, Sandstone, Siltstone, Sediment, Limestone, Dolomite, Coal, Igneous rock, Quartzite and Chert, Anhydrite, Siderite, Redbed. Anhydrite, Siderite, and Redbed are grouped as Others. Figure 3.4 shows an example of data with label extracted by REGEX.



DT	Depth	GR	ILD	ILM	LL8	NPHI	Name	RHOB	SP	Lithology
103.6549	8151.0	32.5818	3.75340	3.05840	4.8650	60.1238	Kugrua I	1.7498	-0.03973	Claystone
62.3583	12403.5	18.9160	634.66022	865.03351	349.7590	0.6110	Kugrua I	2.6681	-12.65312	Sandstone
78.2007	10837.5	80.8075	20.82410	20.69270	21.1136	19.3519	Kugrua I	2.2963	-1.21218	Siltstone
75.6902	10639.5	69.2841	24.37210	19.02810	22.8247	19.7550	Kugrua I	2.0632	-2.97434	Shale
81.9907	9507.5	63.7125	11.61360	12.02750	12.1612	12.6679	Kugrua I	2.5700	-1.98053	Dolomite
NaN	12586.0	NaN	200.56580	89.05580	86.6657	-3.3180	Kugrua I	2.6618	-0.24695	Limestone
78.1815	11198.0	72.6874	18.26400	17.37970	29.4214	10.8962	Kugrua I	2.6432	-0.78205	RedBed
66.8020	11138.5	29.6169	105.84920	66.89490	37.2454	10.3019	Kugrua I	2.7540	-11.76714	RedBeds

Figure 3.4 The rock types which are extracted by REGEX are in the Lithology column

3.2 Data Preparation

Well logging data from 11 wells are split into 3 sets; training set, test set, and validation set. Training set consists of 9 wells while East Simpson II and Ikpikpuk I wells are reserved as test and validation sets, respectively. The set for each well is selected by considering the amount of data of each rock types in that well. The training set is used to train the model while the validation set is for hyperparameter tuning of the model. The accuracy or performance of the model is evaluated from the test set. All the rock types must appear in every set, so the model can be developed and evaluated properly. All log measurements are not measured throughout a borehole well, and some of them are missing during some intervals of depth especially around the top and bottom of the well. In other words, at some depth, the gamma-ray log is available, but the resistivity might be missing. It is possible to fill those missing data with statistical methods such as median, mean, and estimate function. Because filling the missing log measurements add biases to the study, it might be better to drop those depths instead.

3.3 Model development

3.3.1 Exploratory data analysis

Exploratory data analysis is a process which analyzes the dataset to investigate their characteristics. Summary statistics such as arithmetic mean, standard deviation, quartile, minimum, maximum, skewness, kurtosis, for each well log measurement of each label are calculated. The rocks have wide ranges of variation, so well log measurements are overlapped. Multiple well log measurements are also equivalenced to multivariate data. Because of that, the data are very complex and summary statistics are not the most useful calculation for the lithology classification. The amount of data for each rock type is plotted as a barplot to compare the amount of data between the labels. The names of well in which each label appears are listed to be taken into consideration in the training-test-validation set splitting step. Because some intervals of depth inside the well are not interested and are not logged or measured, those missing data are also needed to be counted. Lastly, the statistical correlation is calculated to inspect a relationship between 2 well log measurements.

3.3.2 Building basic models

Next, the basic models are developed by using XgBoost and Random Forest algorithms without any feature engineering or further preprocessing. Performance of the basic models is used as a baseline for accuracy improvement and build advanced models. Since some rock types have a small amount of data, the model has a tendency to ignore them to optimize for homogeneity index. The test data that belong to these minority classes are thus misclassified. Aside from the small amount of data, some labels are not commonly interpreted in traditional well log interpretation such as Argillite and Red beds. This study focuses on 4 main rock types; shale-mudstone, sandstone, limestone, and coal.

3.3.3 Building advanced models

Advanced models are built on basic models by incorporating several feature engineering and preprocessing steps such as oversampling and undersampling, and M-N crossplot. Generally, feature engineering is a process to create new features from the original features by using mathematical operations such as multiplication and division. Ideas for creating the new features are usually derived from knowledge or understanding in the data, but it might be the result of the mathematical operations alone and does not have any real meaning.

3.3.3.1 Oversampling and undersampling

The labels are unbalanced, showing a large difference in the amount of data in each well. Unbalanced data is a common and inevitable problem in machine learning and can be addressed by oversampling or undersampling technique. Undersampling is a sampling technique which subsamples large-amount-of-data labels. In other words, some data from those labels are removed from the training set. Since the data is large enough, they could be treated as a population. For samples to represent the distribution of the population, the appropriate sample size is determined by Yamane's simplified formula (1967) which is defined as;

$$n = \frac{N}{1 + Ne^2}$$

, where n is the sample size, N is the population size, and e is the level of precision. A level of precision is the possible deviation from the sample mean to the population mean and has ranged from 0% to 100%. Also, the formula assumes a 95% confidence level which is the probability that the level of precision holds true. On the other hand, oversampling is a sampling technique which repeatedly selects samples from small-amount-of-data labels. In other words, data of those labels in the training set are duplicate. This technique does not give new insight into the model but it emphasizes the importance of those labels.

3.3.3.2 MN crossplot

A new feature that can be engineered based on data in this study is called MN crossplot. It is a technique which uses density, neutron, and sonic log to identify the well-logging lithology. M and N are defined as follows (Burke et al., 1969);

$$M = \frac{S_{fluid} - S_{log}}{D_{bulk} - D_{fluid}}$$

$$N = \frac{N_{fluid} - N_{log}}{D_{bulk} - D_{fluid}}$$

, where S_{fluid} is the compressional sonic of the drilling mud, S_{log} is the compressional sonic log, D_{bulk} is the density log, D_{fluid} is the density of drilling mud, N_{fluid} is the neutron of the drilling mud, and N_{log} is the neutron log.

3.3.3.3 Scaling and outlier processing

Since gamma-ray log could differ between different wells, the minimum gamma-ray log value in one well might be more than the maximum gamma-ray log value in another well. To be able to compare between different wells, the value must be scaled. A total of 3 scaling methods are used; standardization, normalization, and ranking. First, standardization is defined as follows;

$$GR_{standardized} = \frac{GR_i - \overline{GR}}{s.d.}$$

, where $GR_{standardized}$ is a standardized gamma-ray value, GR_i is an original gamma-ray value, $\overline{GR}$ is a mean of the gamma-ray, and $s.d.$ is a standard deviation of the gamma-raya.

Second, normalization is defined as follows;

$$GR_{normalized} = \frac{GR_i - GR_{min}}{GR_{max} - GR_{min}}$$

, where $GR_{normalized}$ is a normalized gamma-ray value, GR_i is an original gamma-ray value, GR_{max} is a maximum value of the gamma-ray, GR_{min} is a minimum value of the gamma-ray, Next, ranking is ranked ordinally.

For the feature to be scaled, data which greatly differ from the majority or outliers must be dealt with beforehand because mean, maximum, and minimum are sensitive to them. A study by Dixon and Yuen (1974) proposed 2 methods to process and deal with the outliers. First, trimming is a process in which outliers are dropped out. A percentage to drop the data at the minimum and the maximum ends can be specified and varied. On the other hand, winsorizing does not drop the outliers but replace their value by new minimum and maximum which should be obtained after trimming.

Experiments of the preprocessing and feature engineering are conducted separately (Fig 3.5). Each experiment begins with the implementation of one of the preprocessing or feature engineering before the ensemble tree model is trained. The hyperparameters are tuned on the validation set and treated as independent variables while the performance is observed as a dependent variable. Table 3.1 shows a list of hyperparameters which this study focuses on; subsample Hence, the optimal values for each hyperparameter are determined. After that, the model will be evaluated using the test set to compare the impact of each experiment on the model performance. Finally, the advanced models are developed with combinations of all the preprocessing and feature engineering and the performance is evaluated.

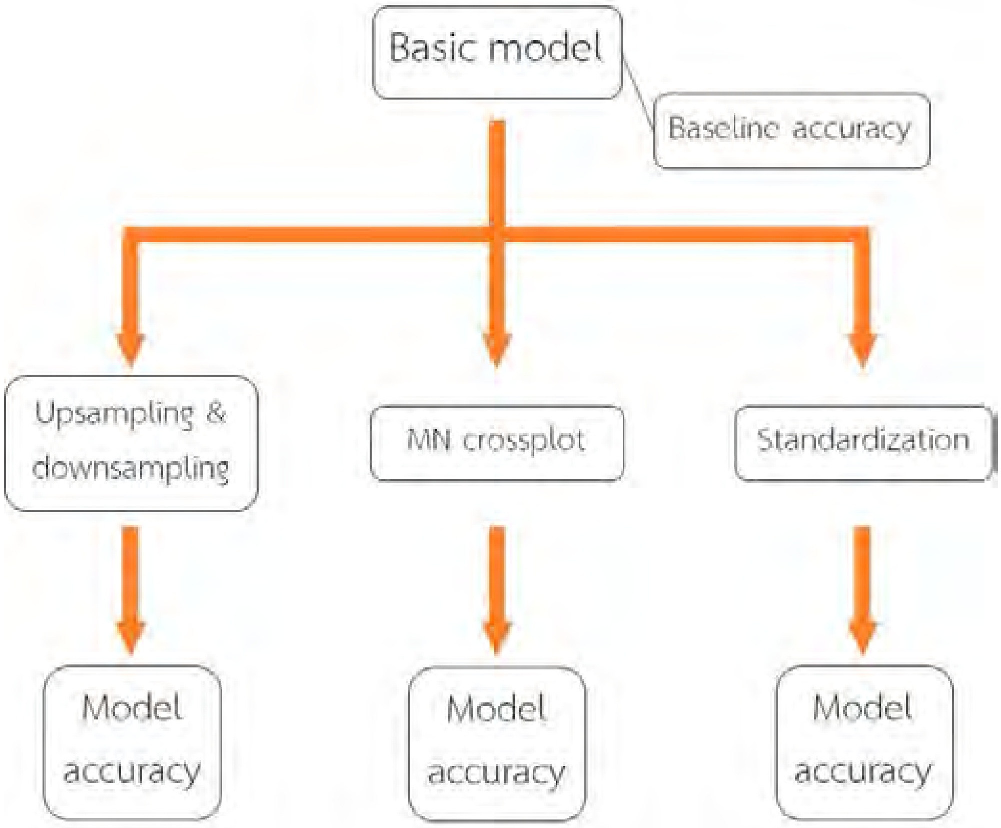


Figure 3.5 Experiments of the preprocessing and feature engineering

Table 3.1 The hyperparameters which are related to ensemble tree model; tree level and ensemble level

	Hyperparameter	Meaning
Tree level	gamma max_depth	A threshold of the homogeneity increase for splitting the node A maximum number of nodes in each decision tree
Ensemble level	learning rate n_estimator	A percentage which each additional tree contribute to the ensemble A number of decision trees to construct as the ensemble

Chapter 4
Results

4.1 Exploratory data analysis

There are a total of 218,038 data points collected from 11 wells. The depth of well ranges from 600 km. to 6,000 km. The types of well-logging include gamma-ray, deep-induction resistivity, medium-induction resistivity, shallow-induction resistivity, neutron, density, sonic, self-potential, caliper, laterolog deep, laterolog shallow. Each well has different types of well-logs but contains 8 logging in common. Well-logging has operated only the depths where it is economics and necessary because it requires resources such as money and time. Hence, some wells contain missing values, and they are removed from the analysis. Missing values account for 40% of the total amount of data. The final amount of data 127,784 well-logging data points are used in this study.

The data contain 6 rock types; mudrock, sandstone, siltstone, limestone, coal, and igneous rocks (Fig 4.1). Mudrock is the most common, and the second most is sandstone. Both of them appear in every well. Coal is the most uncommon, it is so rare that the amount is less than 1% of the mudrock. The rest can only be found in some wells.

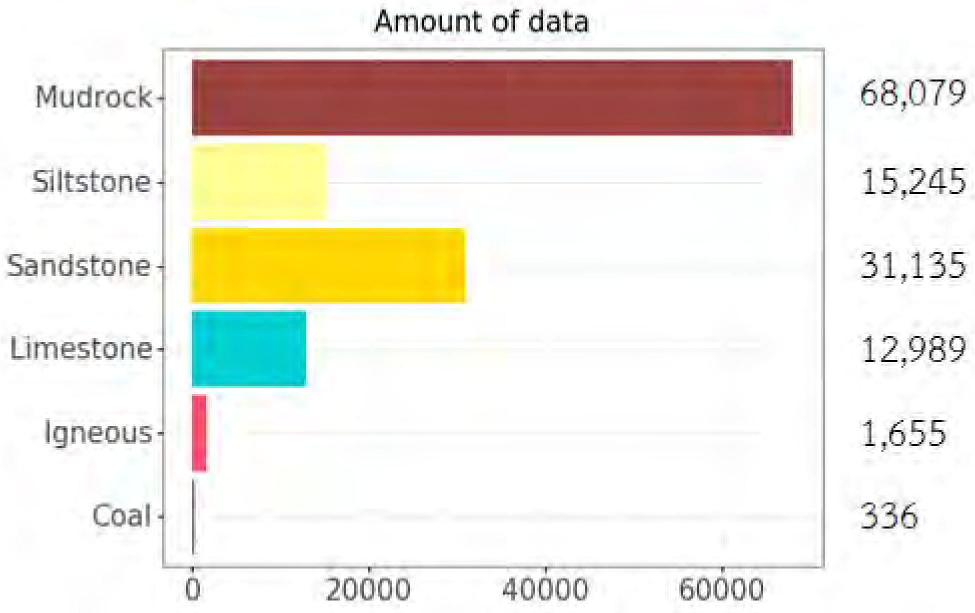


Figure 4.1 Bar chart which visualizes the amount of data in each rock type

Next, to compare gamma and neutron well log measurements between different rock types, a box and whisker plot is used (Fig 4.2). Gamma-ray log measures gamma-ray which is emitted from naturally occurring radioactive content in the rock. Neutron log generates high energy neutron to bombard the rock and the density is estimated from fall off of the neutron. A box and whisker plot shows many aspects of summary statistics (Potter et al., 2010). The box represents 50% of the data and the line inside the box represents the median. The 2 lines which extend from the box allow a possible range for deviation of any datum from the box of majority data. Length of the lines is calculated from the median and interval between the 25th and 75th percentile. Beyond the boundaries, any data are treated as outliers which significantly deviate from the median

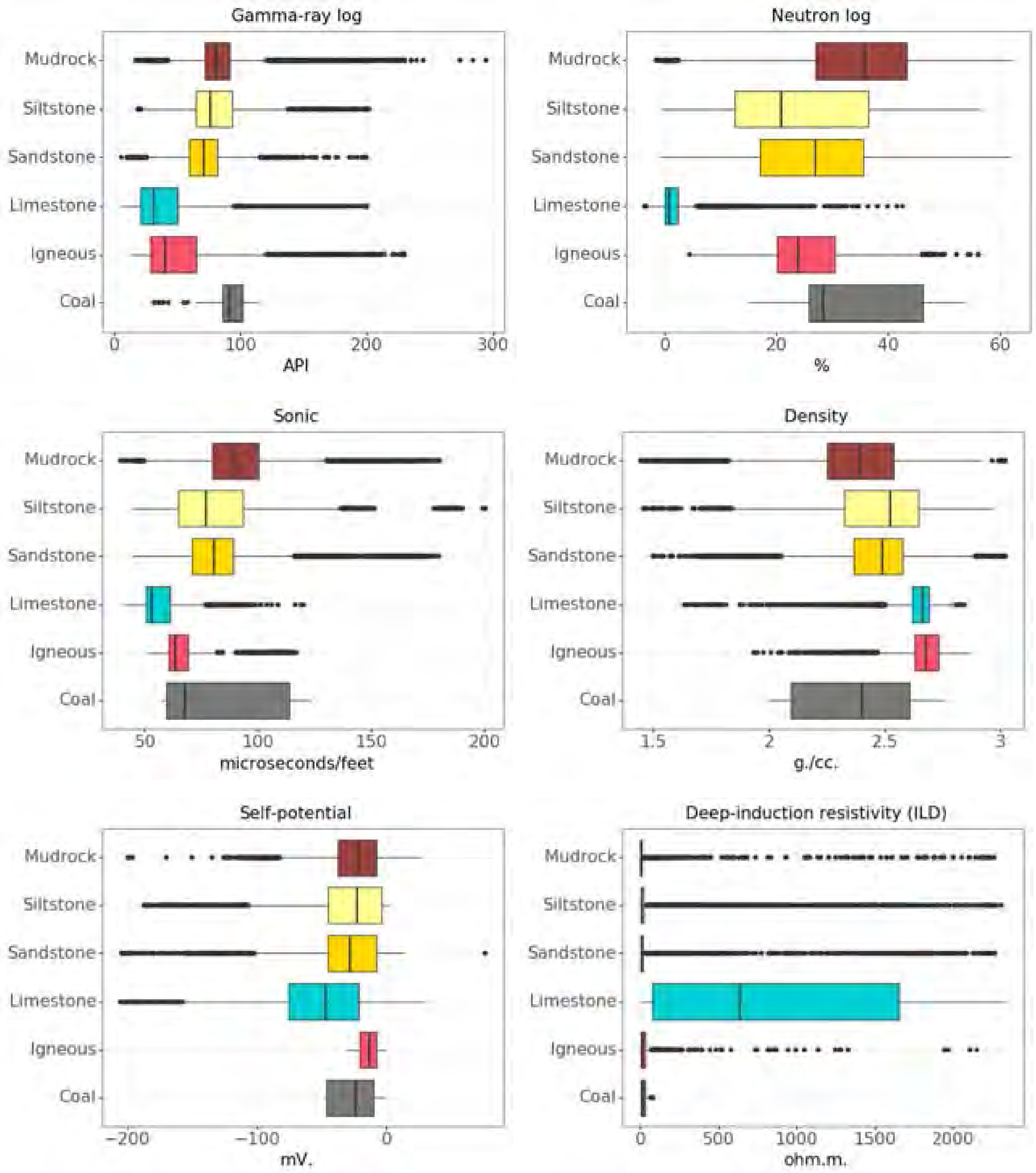


Figure 4.2 The boxplot comparing well log measurement between the rock types. Deep-induction resistivity log is chosen from the 3 resistivity logs which show the same distribution.

Within the scope of gamma-ray and neutron log, limestone has a quite unique characteristic and can be distinguished spontaneously. Igneous rock is also easy to recognize as it has low gamma-ray and high neutron. The rest have similar characteristic which differs only on a small detail i.e. about half of the mudrock has gamma ray higher than all the sandstone. However, taking all logs into consideration at the same time is quite hard, even impossible.

In addition, all pair of well log measurements, the Spearman correlation is visualized in a heatmap (Figure 4.3). The correlation describes a relationship of the pair. The blue indicates a direct proportion between the pair while the red indicates an inverse proportion. The intensity of color describes the strength of the relationship. Certainly, the 3 resistivity logs have a strong direct relationship with each other. Next, there is a strong inverse relationship between depth and two of the density logs, sonic and neutron, because the density of rock tends to increase with depth. However, the density logs itself show a moderate direct relationship with the depth. The rest have a rather weak relationship.

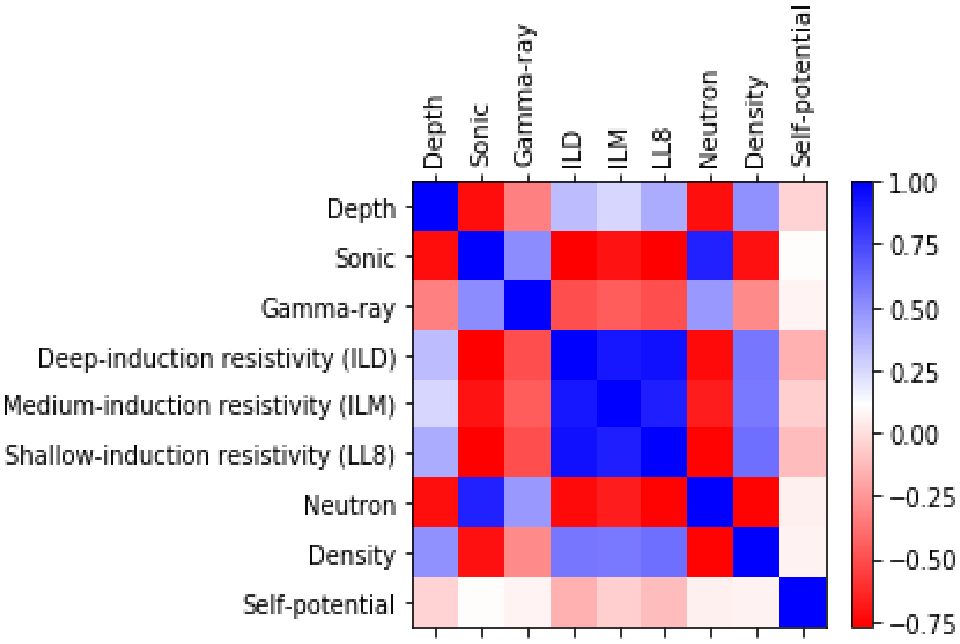


Figure 4.3 Heatmap of the Spearman correlation between well log measurements

4.2 Development of basic model

In order to develop machine learning model for lithology classification, Ikpikpuk I well is treated as the validation data for optimizing the model's hyperparameter and East Simpson II well is treated as the test data for evaluating accuracy or performance of the model. Random forest and Extreme gradient boosting (XGBoost), which are ensemble tree models that have shown outstanding prediction performance by several studies (e.g. Wainer, 2016; Hall and Hall, 2017), is used in this study. Sonic, gamma-ray, deep-induction resistivity, neutron, density, self-potential logs are used as independent variables for the ensemble models.

However, igneous rock is only in 2 wells, so it is impossible to split into 3 sets. Also, well log interpretation rarely interpret rock as siltstone, so siltstone is ambiguous on how it is different from sandstone and mudstone. Because of that, those 2 rock types are excluded from the analysis. The Random forest has random subsampling process and is a stochastic model, therefore to attain the performance of this model with a 95% confidence level, at most 395 rounds of evaluation are needed (Cochran, 1977). Figure 4.4 shows the accuracy of the models. In the XGBoost, 73.6% for East Simpson II, 78.2% for Ikpikpuk I, 77.1% for Inigok I, 58.6% for Kugrua I, 77.3% for South Meade I, and 50.6% for Tunalik I. It is clearly shown that the XGBoost is better than the Random forest and the latter also has inconsistent performance as a result of its stochastic character. Therefore, this study uses only XGBoost model.

Aside from accuracy, the performance of the XGBoost is evaluated with precision, recall, and F1 with respect to each rock types. Table 4.1 - 4.6 show the evaluated. The average F1-score is 36.7% for East Simpson II, 54.3% for Ikpikpuk I, 61.4% for Inigok I, 60.4% for Kugrua I, 43.4% for South Meade I, and 44.9% for Tunalik I. In the situation when the model classifies any data as one rock type, the precision of that rock type reports the chance that this classification is correct. However, the model could attain a high level of precision by classifying only obvious data and avoiding ambiguous data. Because of that, precision alone might not point out a good model, so recall and F1 are used. If the model tries to avoid ambiguous data of any rock type, the recall of that rock type will be low because the model is failed to find all the data in that rock type. Lastly, F1-score is an average score of precision and recall. Afterward, this study attempts to improve the model performance which is measured by these metrics.

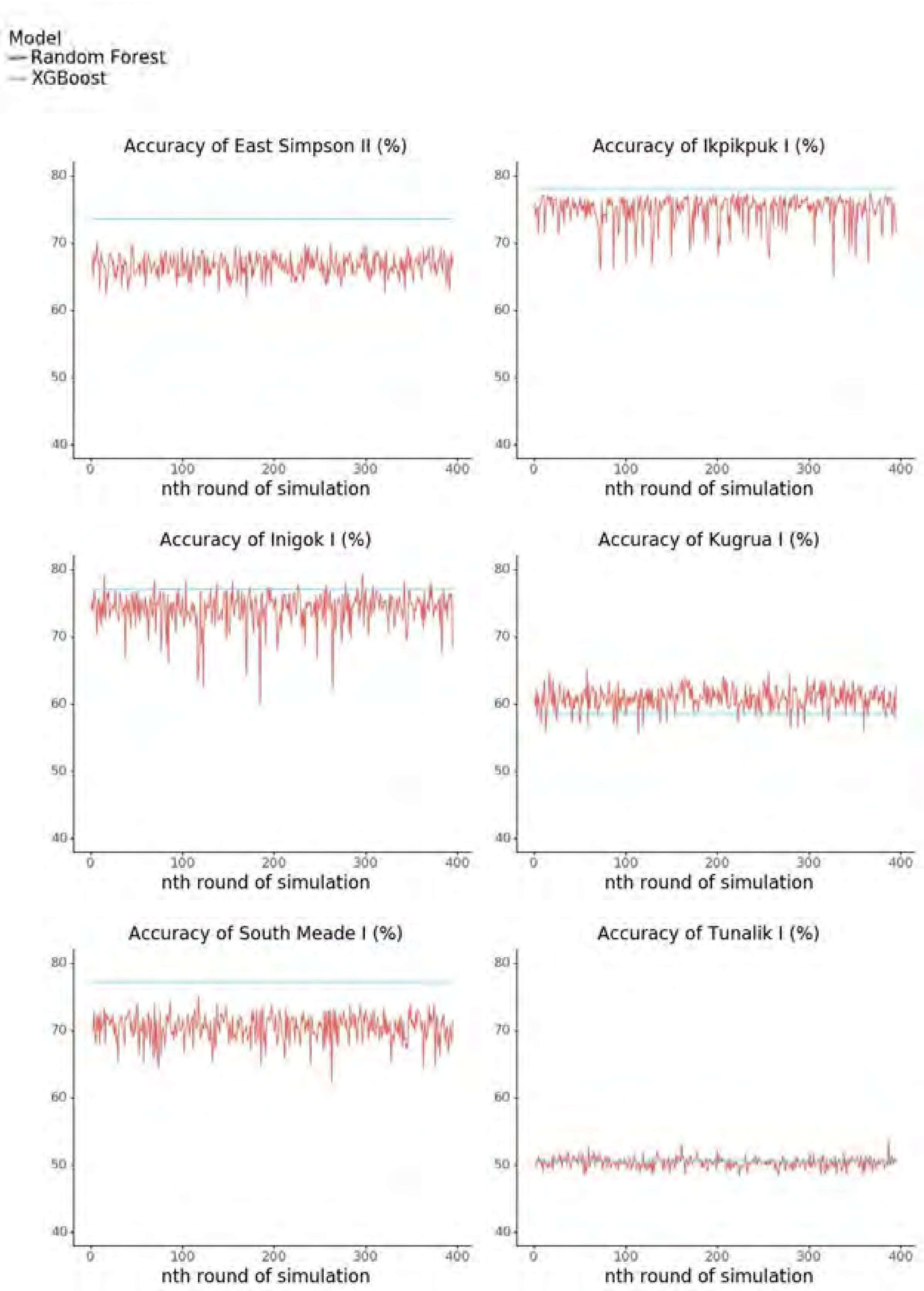


Figure 4.4 Performance of the ensemble models in East Simpson II, Ikpikpuk I, Inigok I, Kugrua I, South Meade I, and Tunalik I

Table 4.1 Detailed evaluation of the baseline performance of the XGBoost in East Simpson II

	Precision	Recall	F1-Score	Amount of data
Mudrock	77.3%	94.9%	85.2%	9,005
Sandstone	39.6%	16.6%	23.4%	2,934
Limestone	98.9%	23.5%	38.0%	374
Coal	0%	0%	0%	76

Table 4.2 Detailed evaluation of the baseline performance of the XGBoost in Ikpikpuk I

	Precision	Recall	F1-Score	Amount of data
Mudrock	77.1%	91.1%	83.5%	11,512
Sandstone	57.5%	34.0%	42.8%	4,749
Limestone	91.4%	90.4%	90.9%	5,566
Coal	0%	0%	0%	100

Table 4.3 Detailed evaluation of the baseline performance of the XGBoost in Inigok I

	Precision	Recall	F1-Score	Amount of data
Mudrock	60.3%	31.3%	41.2%	872
Sandstone	38.3%	76.1%	51.0%	715
Limestone	95.8%	88.3%	91.9%	3,651
Coal	0%	0%	0%	0

Table 4.4 Detailed evaluation of the baseline performance of the XGBoost in Kugrua I

	Precision	Recall	F1-Score	Amount of data
Mudrock	51.0%	83.7%	63.4%	4,767
Sandstone	59.2%	25.5%	35.6%	5,790
Limestone	78.8%	85.6%	82.1%	2,663
Coal	0%	0%	0%	0

Table 4.5 Detailed evaluation of the baseline performance of the XGBoost in South Meade I

	Precision	Recall	F1-Score	Amount of data
Mudrock	92.6%	78.7%	85.1%	10,413
Sandstone	46.6%	76.0%	57.8%	2,483
Limestone	18.5%	87.5%	30.6%	40
Coal	0%	0%	0%	160

Table 4.6 Detailed evaluation of the baseline performance of the XGBoost in Tunalik I

	Precision	Recall	F1-Score	Amount of data
Mudrock	42.8%	87.8%	57.5%	714
Sandstone	0%	0%	0%	825
Limestone	76.2%	78.2%	77.2%	547
Coal	0%	0%	0%	0

Table 4.1 - 4.6 show 0% precision and recall of coal, which is very poor. Table 4.7 shows accuracy and average F1 of the performance evaluation when coal rock type is taken out from the training and test data which shows slightly better performance than before.. Table 4.8 - 4.13 show detailed evaluation without coal rock type.

Table 4.7 Overview of the baseline performance without coal rock type

Test well name	Accuracy	Average F1
East Simpson II	73.8%	48.8%
Ikpikpuk I	78.2%	72.3%
Inigok I	77.0%	61.5%
Kugura I	59.0%	61.0%
South Meade I	78.3%	58.1%
Tunalik I	50.23%	44.8%

Table 4.8 The evaluation of the XGBoost in East Simpson II without Coal

	Precision	Recall	F1-Score	Truth
Mudrock	77.5%	95.0%	85.3%	9,005
Sandstone	38.2%	15.3%	21.9%	2,934
Limestone	99.0%	24.3%	39.1%	374

Table 4.9 The evaluation of the XGBoost in Ikpikpuk I without Coal

	Precision	Recall	F1-Score	Truth
Mudrock	77.7%	90.6%	83.7%	11,512
Sandstone	55.7%	34.8%	42.8%	4,749
Limestone	91.7%	90.0%	90.6%	5,566

Table 4.10 The evaluation of the XGBoost in Inigok I without Coal

	Precision	Recall	F1-Score	Truth
Mudrock	59.7%	32.1%	41.8%	872
Sandstone	38.3%	76.1%	51.0%	715
Limestone	95.8%	87.9%	91.7%	3,651

Table 4.11 The evaluation of the XGBoost in Kugrua I without Coal

	Precision	Recall	F1-Score	Truth
Mudrock	50.9%	84.5%	63.5%	4,767
Sandstone	60.0%	26.2%	36.5%	5,790
Limestone	81.2%	84.6%	82.9%	2,663

Table 4.12 The evaluation of the XGBoost in South Meade I without Coal

	Precision	Recall	F1-Score	Truth
Mudrock	94.2%	78.8%	85.8%	10,413
Sandstone	46.7%	76.0%	57.9%	2,483
Limestone	18.6%	87.5%	30.7%	40

Table 4.13 The evaluation of the XGBoost in Tunalik I without Coal

	Precision	Recall	F1-Score	Truth
Mudrock	42.7%	87.1%	57.3%	714
Sandstone	0.04%	0.03%	0.01%	825
Limestone	75.7%	77.3%	76.5%	547

4.3 Development of advanced model

4.3.1 Upsampling and downsampling

Coal is upsampled and downsampled in the training data, then the model is tested and evaluated. Since both of the processes are related to random sample and stochastic, to attain the performance of this model with a 95% confidence level, at most 395 rounds of evaluation are needed (Cochran, 1977). The Average F1 of baseline, upsampled and downsampled model are compared in Figure 4.5. The effect of downsampling on model performance is inconclusive. Downsampling is good in East Simpson II, and about the same in Ikpikpuk I, while the model performance in South Meade I is worse when the model is downsampled. On the other hand, the performance of the upsampled models are on par with the baseline and even worse in East Simpson II and South Meade I. Afterword, coal rock type is ignored, and South Meade I is eliminate from the the test well name because it doesn’t contain all 3 rock types

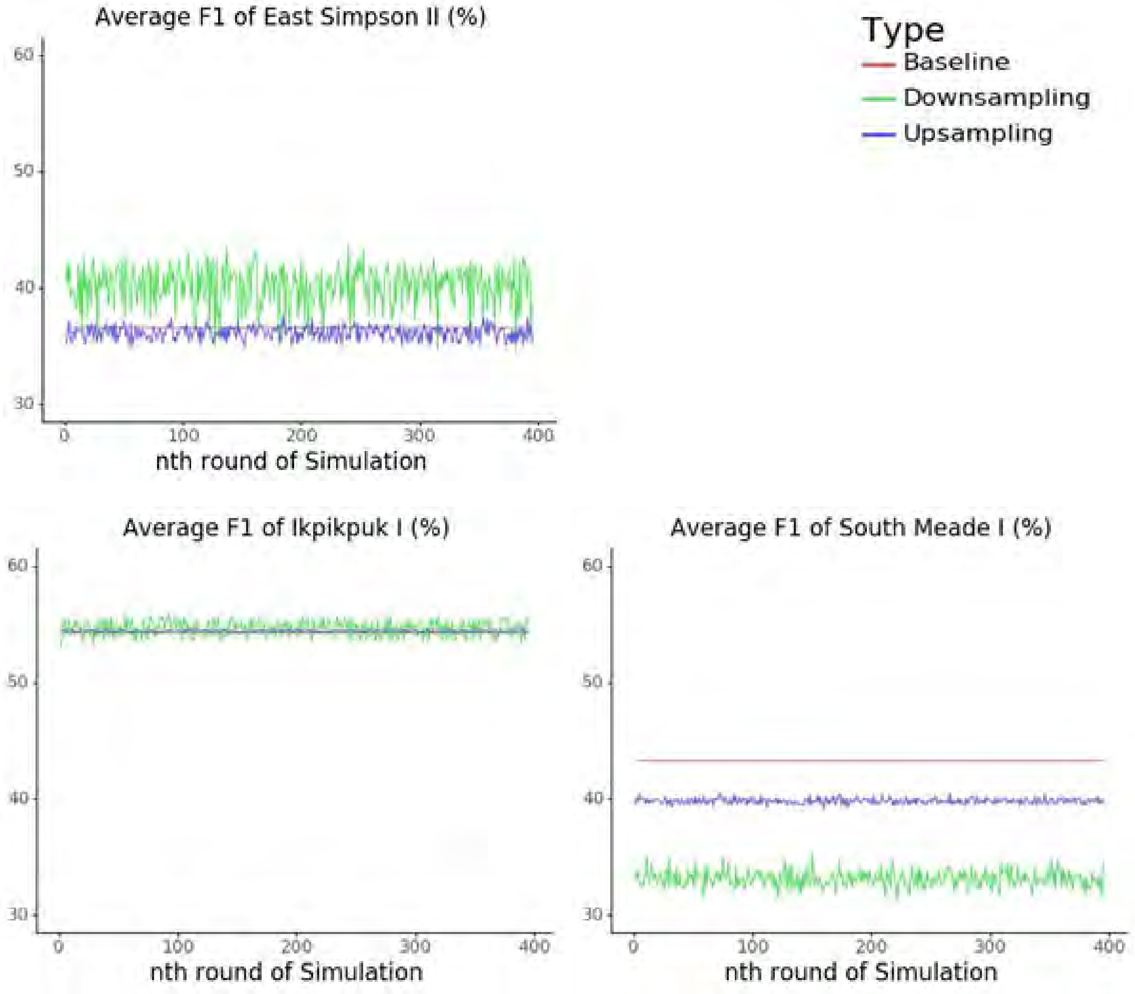


Figure 4.5 Comparison of baseline, upsampled and downsampled model on average F1 score

4.3.2 MN crossplot

M and N are calculated from sonic, neutron, and density logs in both the training and test data. The box and whisker plot is used to compare M and N between different rock types (Fig 4.6). The model performance, both accuracy and average F1, is shown in Figure 4.7. The M and N have a bad effect on a few tests, but they have stronger good influence on some test well. On average, the accuracy is increased by 0.4% and the average F1 is increased by 0.1%.

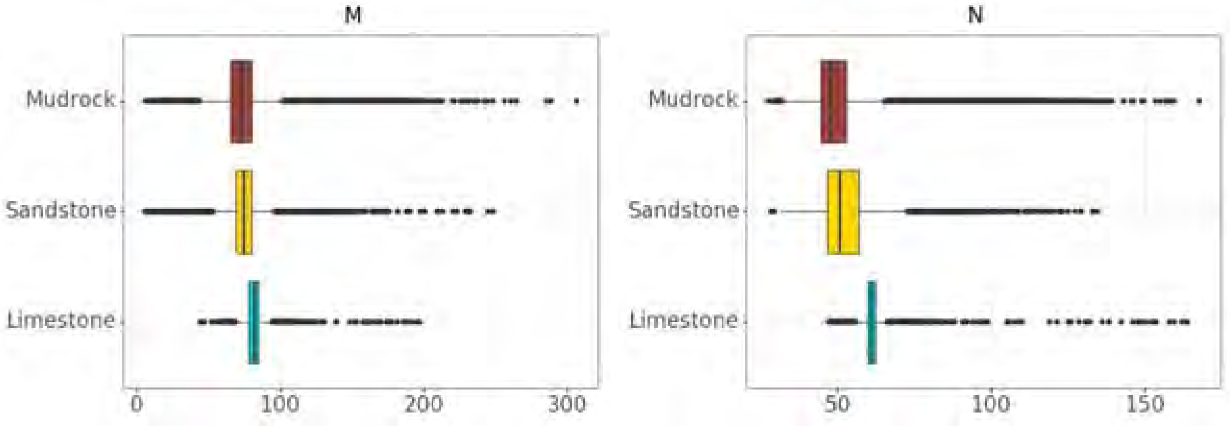


Figure 4.6 The boxplot comparing M and N between the rock types

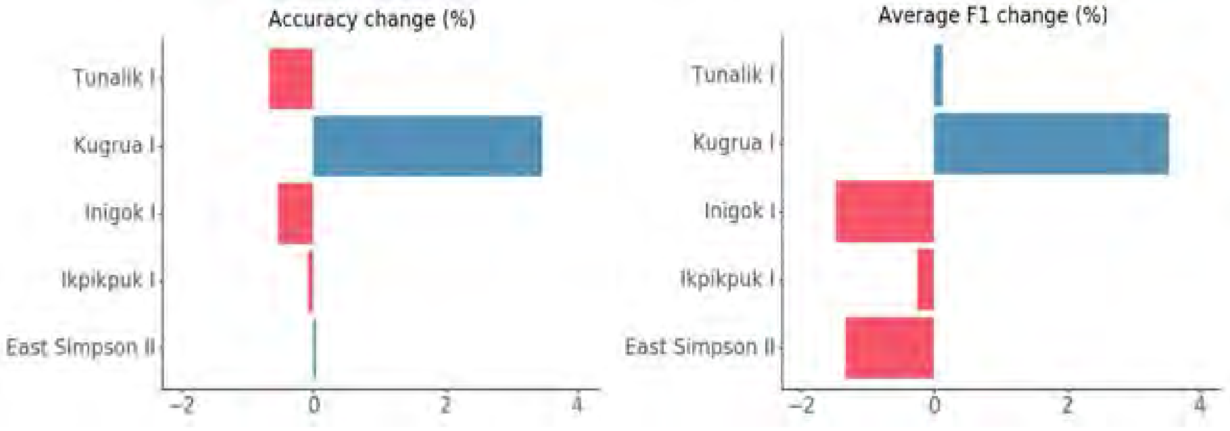


Fig 4.7 Performance change in all 5 test wells after the M and N are added.

4.3.3 Rescaling and Outlier processing

Gamma-ray log is rescaled with all 3 methods: ranking, standardization, normalization, but first the outliers and extreme data are handled with trimming and winsorizing. Percentage of the data which are handled on both minimum and maximum extreme ends are needed to be determined. A various number of percentage are simulated then the model performances are shown in Figure 4.8 for trimming and Figure 4.9 for winsorizing. Afterward, 5% and 10% are selected to use further because the percentage which is more than 10% has a risk to get lower performance than before.

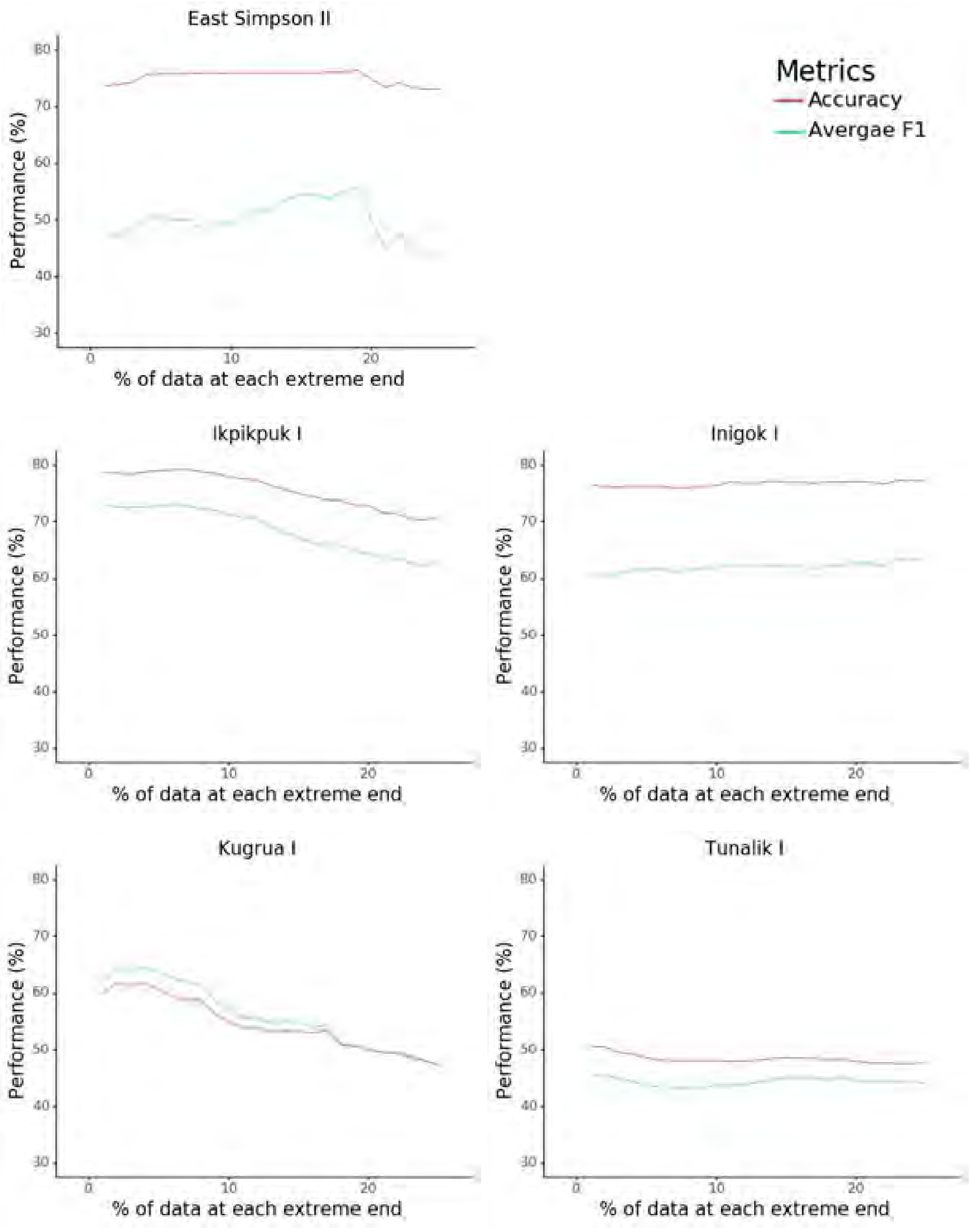


Figure 4.8 Model performances which are trimmed at a various percentage of the data at each extreme end.

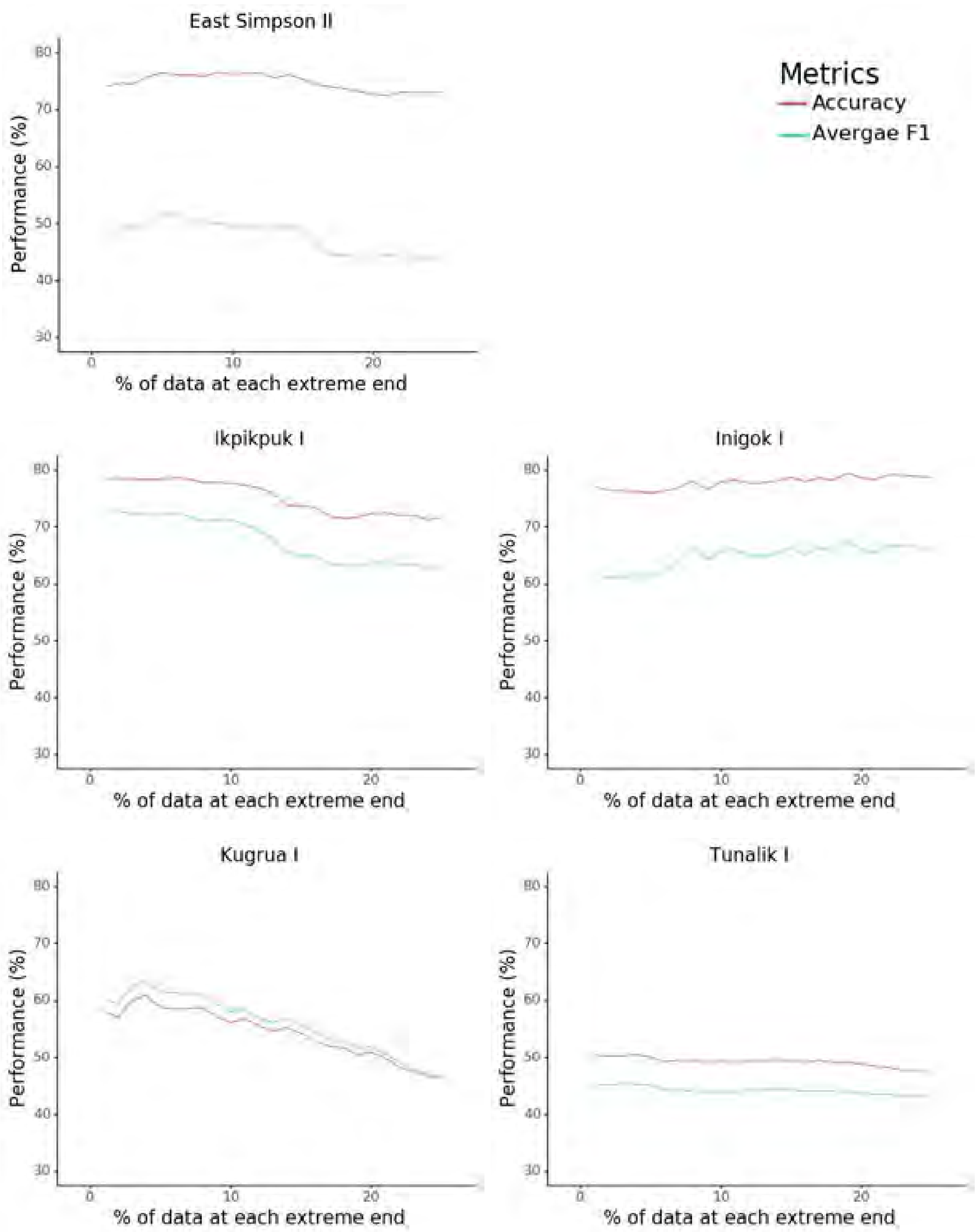


Figure 4.9 Model performances which are trimmed at a various percentage of the data at each extreme end.

Performance of 12 different models is averaged and shown in Figure 4.10. The models are combinations of 2 outlier-handling methods, 2 percentage of handled data, and 3 scaling methods. The useful models are those on the top-right quadrant which are listed in Table 4.14.

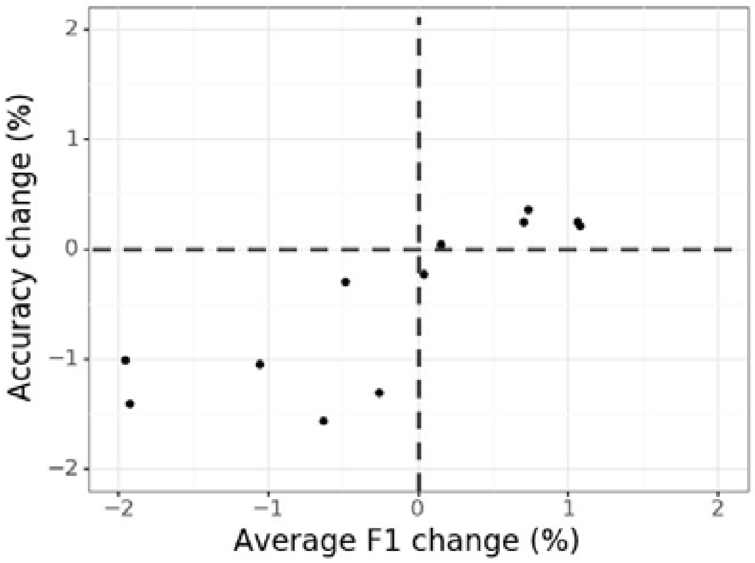


Figure 4.10 Performance of 12 different models from combining outlier-handling methods, percentage of handled data, scaling method

Table 4.14 Detail of the 5 models on the top right quadrant

Combination of methods	Accuracy change (%)	Average F1 change (%)
Trimmed 5% + Normalization	0.24	1.1
Winsorized 5% + Normalization	0.21	1.1
Trimmed 5% + Standardization	0.36	0.74
Winsorized 5% + Standardization	0.24	0.71
Trimmed 5% + Ranking	0.04	0.16

The model which is trimmed 5% at minimum and maximum ends and normalized is selected because it has good performances in both the accuracy and average F1.

4.4 Final model

After experimenting with the M and N indexes, outlier-handling, scaling, to conclude the best performance possible for the model, hyperparameters of the model need to be tuned. The target hyperparameters are gamma, max_depth, learning_rate, and n_estimators which are shown in Table 3.1. The model performance with respect to each of the hyperparameter has experimented separately; Figure 4.11 for gamma, Figure 4.12 for max_depth, Figure 4.13 for learning_rate, and Figure 4.14 for n_estimators. These experiments provide an optimal range of values for each of the hyperparameter.

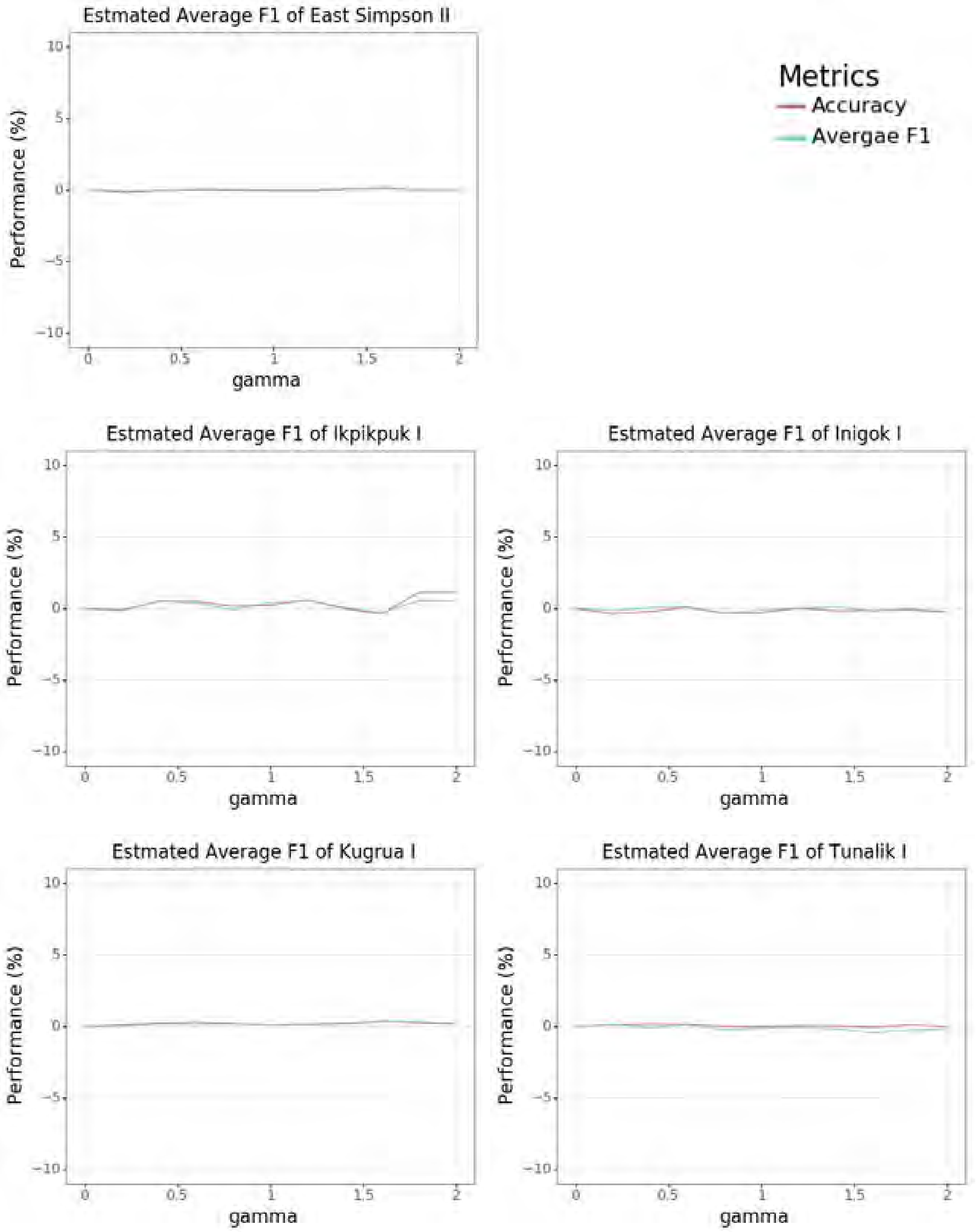


Figure 4.11 The model performance with respect to gamma

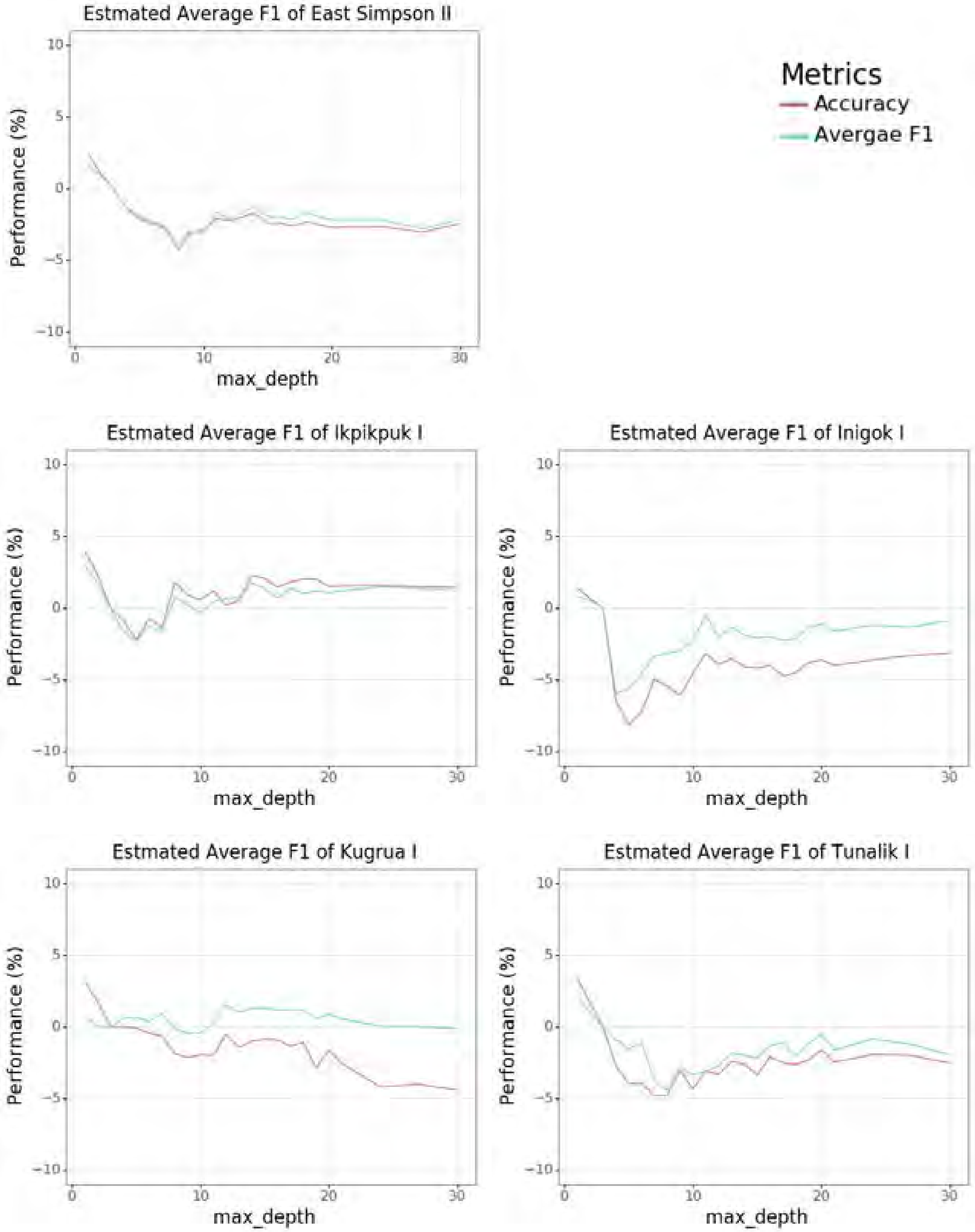


Figure 4.12 The model performance with respect to max_depth

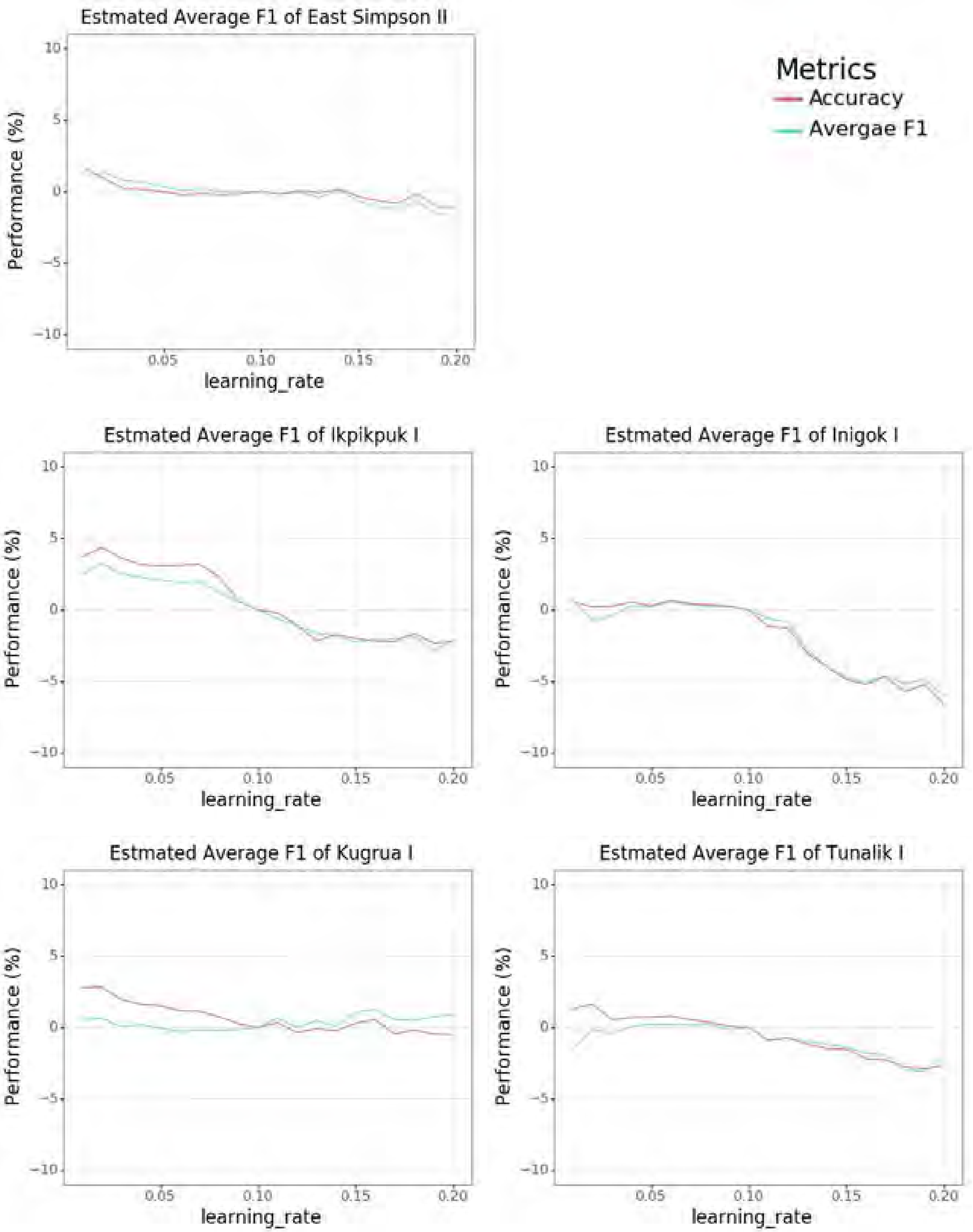


Figure 4.13 The model performance with respect to learning_rate

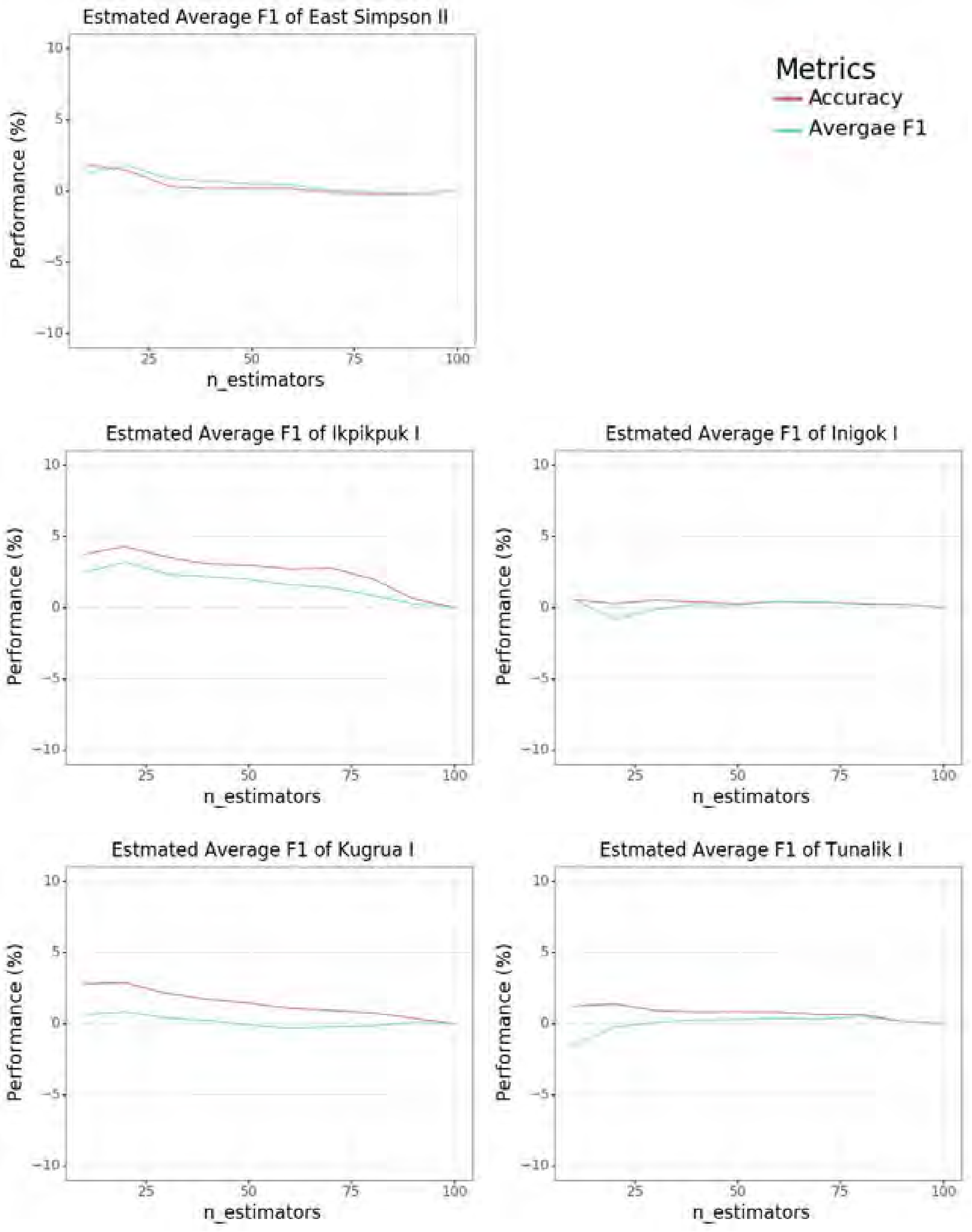


Figure 4.14 The model performance with respect to n_estimators

Next, by selecting one value from the optimal range of each of the hyperparameters at a time and combining them with one selected from the other hyperparameter, various combinations of the hyperparameters are created then all of the hyperparameters are tuned together. Figure 4.15 shows model performances of 341 models from hyperparameter combinations.

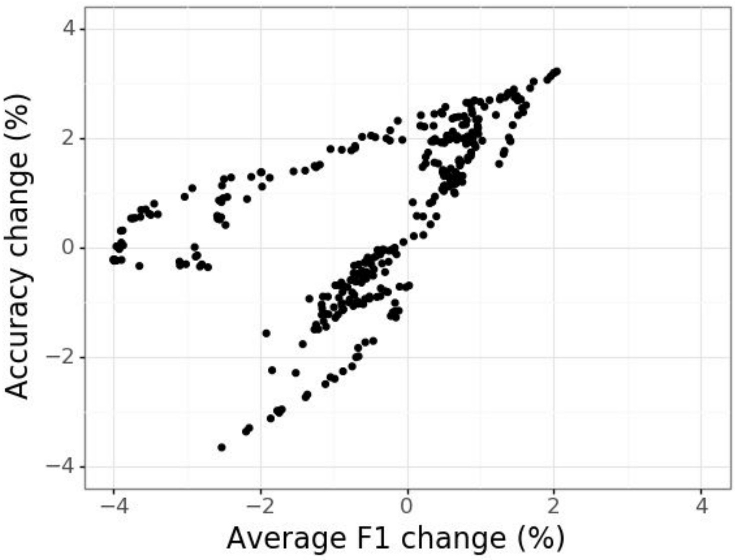


Figure 4.15 The model performance of 341 models from combinations of hyperparameters

The model performances are divided into 2 groups. First, a bottom-middle group which has negative changes in both accuracy and average F1 at the same amount. Another group is a top-left group which has a positive change in accuracy but a stronger negative change in average F1 than the former group. Eventually, the model performances are converged to the top-right group which has positive changes in both accuracy and average F1. Table 4.15 shows 5 combinations of hyperparameters of the most top-right models.

Table 4.15 The hyperparameters of the most top-right models

max_depth	learning_rate	n_estimators	Accuracy change (%)	Average F1 change (%)
1	0.07	110	3.23	2.04
1	0.09	90	3.20	2.01
1	0.07	120	3.20	2.00
1	0.07	130	3.14	1.96
1	0.09	100	3.07	1.91

The best hyperparameters combination is used as the best performance possible for the model. The data is then split into 2 set; training data and test data. The final model is then developed and evaluated (Fig 4.16 and Table 4.16).

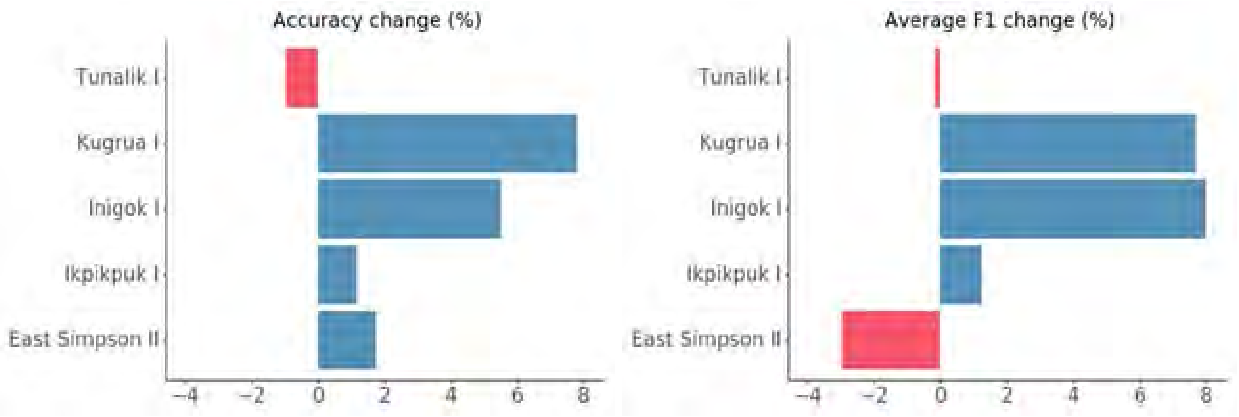


Figure 4.16 The final model performance

Table 4.16 Overview of performances of the model

Test well name	Accuracy	Average F1
East Simpson II	75.6%	45.8%
Ikpikpuk I	79.4%	73.6%
Inigok I	82.5%	69.5%
Kugura I	66.8%	68.6%
Tunalik I	49.3%	44.7%

Table 4.17 The evaluation of the XGBoost in East Simpson II

	Precision	Recall	F1-Score	Truth
Mudrock	78.4%	96.7%	86.6%	9,005
Sandstone	48.0%	19.0%	27.2%	2,934
Limestone	98.0%	13.4%	23.5%	374

Table 4.18 The evaluation of the XGBoost in Ikpikpuk I

	Precision	Recall	F1-Score	Truth
Mudrock	78.8%	91.3%	84.6%	11,512
Sandstone	59.8%	36.1%	45.0%	4,749
Limestone	90.7%	91.7%	91.1%	5,566

Table 4.19 The evaluation of the XGBoost in Inigok I

	Precision	Recall	F1-Score	Truth
Mudrock	69.4%	53.6%	60.5%	872
Sandstone	47.6%	63.1%	54.3%	715
Limestone	94.1%	93.2%	93.6%	3,651

Table 4.20 The evaluation of the XGBoost in Kugrua I

	Precision	Recall	F1-Score	Truth
Mudrock	61.3%	78.8%	69.0%	4,767
Sandstone	68.6%	47.6%	56.2%	5,790
Limestone	75.3%	85.0%	80.7%	2,663

Table 4.21 The evaluation of the XGBoost in Tunalik I

	Precision	Recall	F1-Score	Truth
Mudrock	42.3%	83.6%	56.2%	714
Sandstone	0.15%	0.02%	0.03%	825
Limestone	72.9%	75.9%	74.4%	547

Chapter 5
Discussion and conclusion

5.1 Discussion

5.1.1 Dataset and model performance

Table 5.1 compares the number of classes to classify, the used model, and the final average F1 between this study and others (Chen and Zeng, 2018; Messer et al., 2017; Hall, 2016). Those 3 studies used the same dataset. which has 9 classes; non-marine sandstone, non-marine coarse siltstone, non-marine fine siltstone, marine siltstone and shale, mudstone, wackestone, dolomite, packstone-grainstone, and phylloid-algal bafflestone. The studies achieve about the same model performance as this study even though they need to classify more classes. This might be due to 2 advantages of their studies; geologist-defined feature and depth interval of the data. First, their studies use various features which including common logs such as gamma-ray, resistivity, phi neutron-density, delta neutron-density, and photoelectric. However, some additional features are created by a geologist such as marine/nonmarine indicator and relative position are provided, and these features are useful for classifying their 9 classes of rocks. For another advantage, the data in their studies are from about the same depth, so the heterogeneity might not be as strong as in our study.

Table 5.1 Comparison between this study and others on the number of classes to classify, the used model, and the F1

	Number of classes	Model	F1 (%)
This study	3	XGBoost	60.5
Chen and Zeng (2018)	9	XGBoost	61.0
Mosser et al. (2017)	9	XGBoost	58.0
Hall (2016)	9	Support Vector Machine	43.0

5.1.2 Hyperparameters

Table 5.2 compares hyperparameters which are focused in this study with Mosser et al. (2017) which is the only study that published the hyperparameters among those 3 studies. Generally, the model with high complexity is required for a large number of classes to classify and complexity of the XGBoost is partly controlled by max_depth and n_estimators. Complexity of the XGBoost is directly proportional to max_depth and n_estimators. Thus, due to the number of classes, the model in a study by Mosser et al. (2017) is more complex than the model in this study while both studies reaches the same degree of performances.

Table 5.2 Comparison between this study and Mosser et al. (2017) on the complexity

	This study	Mosser et al. (2017)
Number of classes	3	9
max_depth	1	3
n_estimators	110	150

5.1.3 Features engineering

Out of 4 data preprocessing and feature engineering, upsampling and downsampling is the only one which is not working well. A study by Ling and Li (1998) demonstrates the effect of upsampling and downsampling on model performance with business data. The upsampling method doesn’t improve the model performance while the downsampling method significantly improves the model performance. Furthermore, a study by Japkowics (2000) also simulates the effect of upsampling and downsampling on synthetic data. As a result, both of them improve model performance. Because of that, the downsampling might be a useful preprocessing method even though it is inconclusive in this study.

5.2 Further study

Machine learning considers each row of data separately. However, because the rock is thick, well logging data is spatially correlated and should not be considered separately. This fact might be incorporated to improve the model. Moreover, in general, well log interpretation is done on an economic interval of depth or a specific formation. Therefore, the study can focus on the specific interval to achieve better performance and to respond with a use case.

5.3 Conclusion

In this study, an ensemble tree which is called extreme gradient boosting is used to classify rock types from well-logging data. The target rock types are mudrock, sandstone, and limestone. The baseline performances are shown in Table 4.7 and the average accuracy and average F1 are 67.6% and 57.7% respectively. Originally, coal is also one of the targets, but it is ignored because of its extremely small amount of data. A total of 4 data preprocessing and feature engineering are demonstrated; upsampling and downsampling, MN crossplot, rescaling, and outlier processing.

First of all, after coal rock type is upsampled in the training data, the performances are worse than the baseline which is shown in almost all of the experimental simulation (Fig 4.5). Thus, the effect of upsampling on model performance is negative. On the other hand, the effect of downsampling on model performance can't be concluded. This is deduct from the experiment in which the training data is downsampled, and the model performances are better than the baseline in some test well and are worse than the baseline in some test well (Fig 4.5).

Second, MN crossplot also leads to both increase and decrease of the performance which is depending on the test well. However, on average, the accuracy and the average F1 are increased by 0.4% and 1.2%, respectively, so the M and N are probably useful for the model.

Next, rescaling and outlier processing. An appropriate percentage cutoff for the outlier processing is determined from the experiments (Fig 4.8 - 4.9). Then, outlier preprocessing methods which are trimming and winsorizing are incorporated with the rescaling methods which are standardization, normalization, and ranking (Fig 4.10). The 2 methods of outlier processing show approximately the same model performance and the most fitting cutoff is 5%. The best rescaling method is normalization. Rescaling and outlier preprocessing averagely improve the accuracy and average F1 by 0.2% and 1.1%.

Afterward, hyperparameters which are focused in this study; max_depth, n_estimators, learning_rate, and gamma are tuned. The most fitting values are 1, 110, 0.07, and 0, respectively. The final model performance is shown in Table 4.16, and on average, the accuracy and average F1 are 70.7% and 60.5% which are improved from the baseline by 3.1% and 2.8% respectively.

List of references

Archie, G. (1942). The Electrical Resistivity Log as an Aid in Determining Some Reservoir Characteristics. *Transactions of the AIME, 146*(01), 54-62.

Barber, D. (n.d.). Machine learning. *Bayesian Reasoning and Machine Learning,* 303-304.

Bestagini, P., Lipari, V., & Tubaro, S. (2017). A machine learning approach to facies classification using well logs. *SEG Technical Program Expanded Abstracts 2017*.

Bird, K. J. (2001). Framework Geology, Petroleum Systems, And Play Concepts Of The National Petroleum Reserve – Alaska. *Petroleum Plays and Systems in the National Petroleum Reserve–Alaska,* 5-17.

Bird, K.J., and Molenaar, C.M., (1987), Stratigraphy, *Petroleum geology of the northern part of the Arctic National Wildlife Refuge, northeastern Alaska: U.S. Geological Survey Bulletin 1778*, 37-59.

Breiman, L. (2001). *Machine Learning, 45*(1), 5-32.

Burke, J., Campbell, R., & Schmidt, A. (1969). The Litho-Porosity Cross Plot A Method Of Determining Rock Characteristics For Computation Of Log Data. *SPE Illinois Basin Regional Meeting*.

Busch, J., Fortney, W., & Berry, L. (1987). Determination of Lithology From Well Logs by Statistical Analysis. *SPE Formation Evaluation, 2*(04), 412-418.

Ceriani, L., & Verme, P. (2011). The origins of the Gini index: Extracts from Variabilita e Mutabilita (1912) by Corrado Gini. *The Journal of Economic Inequality, 10*(3), 421-443.

Chen, J., & Zeng, Y. (2018). 2018, Application of Machine Learning in Rock Facies Classification with Physics-Motivated Feature Augmentation.

Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*.

Cochran, W. G. (1977). Sampling Techniques, 3rd Edition. John Wiley. ISBN: 0-471-16240-X

Dixon, W. J., & Yuen, K. K. (1974). Trimming and winsorization: A review. Statistische Hefte, 15(2-3), 157-170.

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.

Enikanselu P. A., & Ojo A. O. (2012), Statistical analysis and evaluation of lithofacies from wireline logs over ‘Beleema’ field, Niger Delta, Nigeria. *Journal of Petroleum and Gas Engineering,* 3(2), 26-34.

Hall, B. (2016). Facies classification using machine learning. *The Leading Edge, 35*(10), 906-909.

Hall, M., & Hall, B. (2017). Distributed collaborative prediction: Results of the machine learning contest. *The Leading Edge, 36*(3), 267-269.

Howell, D. G., Jones, D. L., & Schermer, E. R. (1983). Tectonostratigraphic terranes of the frontier circum-Pacific region*. AAPG Bulletin*, 67.

Hubbard, R. J., Edrich, S. P., & Rattey, R. P. (1987). Geologic evolution and hydrocarbon habitat of the 'Arctic Alaska Microplate'. *Marine and Petroleum Geology, 4*(1), 2-34.

Japkowicz, N. (2000). The class imbalance problem: Significance and strategies. *In Pro-ceedings of the 2000 International Conference on Artificial Intelligence (IC-AI'2000): Special Track on Inductive Learning.*

Jones, D., Howell, D., Coney, P., & Monger, H. (1983). Recognition, Character and Analysis of Tectonostratigraphic Terranes In Western North America. *Journal of Geological Education, 31*(4), 295-303.

Kent, A., Berry, M. M., Luehrs, F. U., & Perry, J. W. (1955). Machine literature searching VIII. Operational criteria for designing information retrieval systems. *American Documentation, 6*(2), 93-101.

Kotsiantis, S. B. (2011). Decision trees: A recent overview. *Artificial Intelligence Review, 39*(4), 261-283.

Lerand, M., (1973), Beaufort Sea, *The future petroleum provinces of Canada – their geology and potential: Canadian Society of Petroleum Geology Mem-oir 1*, 315-386.

Ling, C., & Li, C. (1998). Data Mining for Direct Marketing Problems and Solutions. In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98)*

Loh, E. (2018). Medicine and the rise of the robots: A qualitative review of recent advances of artificial intelligence in health. *BMJ Leader, 2*(2), 59-63.

Marsh, G.M., & Seo, S. (2006). A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets.

Moore, T. E., Wallace, W. K., Bird, K. J., Karl, S. M., Mull, C. G., & Dillon, J. T. (n.d.). Geology of northern Alaska. *The Geology of Alaska,* 49-140.

Mosser, P., & Briceno, A. (2017). https://github.com/seg/2016-ml-contest/tree/master/LA Team

Morgan, J. N., & Sonquist, J. A. (1963). Problems in the Analysis of Survey Data, and a Proposal. *Journal of the American Statistical Association, 58*(302), 415-434.

Navada, A., Ansari, A. N., Patil, S., & Sonkamble, B. A. (2011). Overview of use of decision tree algorithms in machine learning. *2011 IEEE Control and System Graduate Research Colloquium*.

Potter, K., Kniss, J., Riesenfeld, R., & Johnson, C. (2010). Visualizing Summary Statistics and Uncertainty. *Computer Graphics Forum, 29*(3), 823-832.

Prasad, S., Savithri, T. S., & Krishna, I. V. (2017). Comparison of Accuracy Measures for RS Image

Classification using SVM and ANN Classifiers. *International Journal of Electrical and Computer Engineering (IJECE), 7*(3), 1180.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal, 27*(3), 379-423.

Silberling, N.J., Jones, D.L., Monger, J.W.H., Coney, P.J., Berg, H.C., and Plafker, George. (1994). Lithotectonic terrane map of Alaska and adjacent parts of Canada, 1 sheet, 1:2,500,000. *The Geology of Alaska*.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., . . . Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature, 550*(7676), 354-359.

Sudakov, O., Burnaev, E., & Koroteev, D. (2019). Driving digital rock towards machine learning: Predicting permeability with gradient boosting and deep neural networks. *Computers & Geosciences*.

U.S. Geological Survey, Department of the interior. (1999). *USGS Open File Report 99-015*

U.S. Geological Survey, Department of the interior. (2000). *USGS Open File Report 00-200*

U.S. Geological Survey, Department of the interior. (2010). 2010 Updated Assessment of Undiscovered Oil and Gas Resources of the National Petroleum Reserve in Alaska (NPRA). *Fact Sheet 2010–3102*.

Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F., & Mueller, A. (2015, 06). Scikit-learn. *GetMobile: Mobile Computing and Communications, 19*(1), 29-33.

Wahrhaftig, C. (1965). Physiographic divisions of Alaska. *Professional Paper*.

Wainer, J. (2016). Comparison of 14 different families of classification algorithms on 115 binary datasets. arXiv:1606.00930