

# **FACIES PREDICTION USING MACHINE LEARNING ALGORITHM**

**By**

**OBASEESIN KEHINDE ABIODUN**

**PSC 1707831**

**DEPARTMENT OF PHYSICS  
FACULTY OF PHYSICAL SCIENCES  
UNIVERSITY OF BENIN  
BENIN CITY.**

**January , 2023**

# **FACIES PREDICTION USING MACHINE LEARNING ALGORITHM**

**By**

**OBASEESIN KEHINDE ABIODUN**

**PSC 1707831**

**A THESIS SUBMITTED TO THE DEPARTMENT OF PHYSICS  
FACULTY OF PHYSICAL SCIENCES, UNIVERSITY OF BENIN,  
BENIN CITY.**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR  
THE AWARD OF A BACHELOR OF SCIENCE (B.SC.) DEGREE  
IN  
PHYSICS.**

**January , 2023**

# CERTIFICATION.

This is to certify that this research work was carried out and presented by Mr. KEHINDE OBASESIN. Of the department of physics, faculty of physical sciences, university of Benin City, Edo state, Nigeria.

.....  
**DR IBRAHIM T. TAHIRU**  
Project supervisor

.....  
DATE

.....  
**PROF. O.D OSAHON**  
(Head of Department)

.....  
DATE

.....  
(External Examiner)

.....  
DATE

# DEDICATION

I would like to dedicate this project work to God, who has been my rock and my guide throughout my academic journey. His love and grace have sustained me and have given me the strength and determination to persevere through the challenges and obstacles that I have faced. I am deeply grateful for His presence in my life and for the countless blessings that He has bestowed upon me.

I would also like to dedicate this project work to my loved ones who have supported me along the way. To my parents, who have always believed in me and encouraged me to pursue my dreams, thank you for your love and guidance. To my friends and peers, who have provided me with motivation and inspiration, thank you for your support and companionship. Finally, to my professors and mentors, who have imparted their wisdom and knowledge upon me, thank you for your guidance and encouragement. Your love and support have been invaluable to me and have helped me reach this point in my academic career. I am deeply grateful for all that you have done for me.

## **CERTIFICATION ON DISSERTATION ON PLAGIARISM**

We the undersigned attest and declare that the dissertation OBASESIN KEHINDE titled FACIES PREDICTION USING MACHINE LEARNING ALGORITHM, has passed the anti-plagiarism test and does not violate any copyright regulations.

.....

.....

**DR. IBRAHIM.T.TAHIRU.**

**DATE**

(Project supervisor)

.....

.....

**PROF. O.D OSAHON**

**DATE**

(Head of Department)

## **ACKNOWLEDGEMENTS**

I am deeply grateful to my supervisor, for his guidance and support throughout the course of this project. His expertise and patience have been invaluable to me, and I have greatly benefited from his insights and advice. I am also grateful for the opportunity to work under their supervision and to learn from their wealth of knowledge and experience.

I would also like to thank THE UNIVERSITY OF BENIN for providing me with the resources and support needed to complete this project. I am grateful for the opportunity to be a part of this community and to learn from my fellow students and professors.

Finally, I would like to extend my heartfelt gratitude to my family and friends, who have supported me every step of the way. Your love and encouragement have meant the world to me, and I could not have completed this project without your help. Thank you all for your unwavering support.

## **ABSTRACT.**

Facies prediction refers to the task of determining the type of rock or sediment in a particular area, based on various physical and chemical properties. Machine learning techniques can be used to build predictive models for facies prediction, using data on the characteristics of different rock types and the corresponding measurements made at various locations. These models can then be used to make predictions about the facies of new, previously unseen locations.

There are several benefits to using machine learning for facies prediction. One benefit is that the models can be trained on large amounts of data, allowing them to make highly accurate predictions. Additionally, machine learning models can be updated as new data becomes available, enabling them to improve over time.

There are also several challenges to using machine learning for facies prediction. One challenge is the need to carefully preprocess the data to ensure that it is suitable for use in a predictive model. Another challenge is the need to select an appropriate machine learning algorithm for the task, as different algorithms may be more or less suited to different types of data and prediction tasks.

Despite these challenges, machine learning has the potential to greatly improve facies prediction by providing more accurate and up-to-date predictions than traditional methods. As such, it is an active area of research in the geosciences

# TABLE OF CONTENTS

<b>CONTENTS</b>	
<b>TITLE PAGE</b>	<b>i</b>
<b>CERTIFICATION.</b>	<b>iii</b>
<b>DEDICATION</b>	<b>iv</b>
<b>CERTIFICATION ON DISSERTATION ON PLAGIARISM</b>	<b>v</b>
<b>ACKNOWLEDGEMENTS</b>	<b>vi</b>
<b>ABSTRACT.</b>	<b>vii</b>
<b>TABLE OF CONTENTS</b>	<b>viii</b>
<b>LIST OF FIGURES.</b>	<b>x</b>
<b>LIST OF TABLES</b>	<b>xi</b>
<b>CHAPTER ONE</b>	<b>1</b>
<b>INTRODUCTION AND LITERATURE REVIEW</b>	<b>1</b>
<b>GENERAL OVERVIEW</b>	<b>1</b>
<b>1.1 DEFINITION OF GEOSCIENCE:</b>	<b>1</b>
<b>1.2 GEOLOGY:</b>	<b>2</b>
<b>1.1.1 MINEROLOGY:</b>	<b>2</b>
<b>1.1.2 GEOCHEMISTRY:</b>	<b>2</b>
<b>1.1.3 PETROLOGY:</b>	<b>2</b>
<b>1.1.4 PALEONTOLOGY:</b>	<b>3</b>
<b>1.1.5 STRATIGRAPHY:</b>	<b>3</b>
<b>1.1.6 SEDIMENTOLOGY:</b>	<b>3</b>
<b>1.1.7 GEOMORPHOLOGY:</b>	<b>3</b>
<b>1.1.8 HYDROGEOLOGY:</b>	<b>3</b>
<b>1.1.9 MINING GEOLOGY:</b>	<b>4</b>
<b>1.2 GEOPHYSICS:</b>	<b>4</b>
<b>1.3 GEOPHYSICAL METHODS</b>	<b>4</b>
<b>1.4 PETROLEUM GEOSCIENCE:</b>	<b>11</b>
<b>1.4.1 WELL LOGGING; DEFINITION, TYPES AND USES</b>	<b>12</b>
<b>1.4.2 EVOLUTION OF WELL LOGGING.</b>	<b>12</b>
<b>1.4.3 TYPES OF WELL LOGGING</b>	<b>13</b>

1.4.4	<b>AIM OF STUDY</b>	16
1.4.5	<b>OBJECTIVES OF STUDY:</b>	16
	<b>CHAPTER 2</b>	18
	<b>LITERATURE REVIEW.</b>	18
	<b>MACHINE LEARNING IN GEOSCIENCE</b>	18
2.1	<b>HISTORY OF MACHINE LEARNING;</b>	18
2.2	<b>SUPERVISED LEARNING:</b>	19
2.3	<b>UNSUPERVISED LEARNING:</b>	19
2.3	<b>REINFORCEMENT LEARNING:</b>	19
2.4	<b>HISTORY OF MACHINE LEARNING APPLICATION IN GEOPHYSICS</b>	
	20	
2.5	<b>Machine learning algorithms. (Classification).</b>	21
	<b>CHAPTER 3</b>	29
	<b>METHODOLOGY</b>	29
3.1.	<b>MATERIALS AND METHOD:</b>	29
3.2.	<b>METHOD</b>	29
	<b>CHAPTER 4</b>	33
	<b>RESULT AND DISCUSSION</b>	33
4.1	<b>RESULTS:</b>	33
	<b>CHAPTER 5</b>	47
	<b>FINDINGS, CONCLUSION, AND SUGGESTION FOR FURTHER STUDIES.</b>	
	47	
5.1	<b>FINDINGS:</b>	47
5.2	<b>CONCLUSION</b>	47
5.3	<b>SUGGESTION FOR FURTHER STUDIES:</b>	48
	<b>REFERENCES</b>	49

## **LIST OF FIGURES.**

4.1 Total facie curve for well 1

4.2 Comparison between actual and predicted facie for well 1

4.3 Total facie curve for well 2.

4.4 Comparison between actual and predicted facie for well 2

4.5 Total facie curve for well 3

4.6 Total facie curve for well 4

4.7 Comparison between the actual and predicted facie for well 3 and 4.

4.8 Predicted facie for well 5

4.9 Total facie curve for well 5

## **LIST OF TABLES**

1.1 Geographical survey method and their related parameters

3.2 Checking variables in all wells

4.1 Table showing all the used machine learning algorithms and their performance.

# **CHAPTER ONE**

## **INTRODUCTION AND LITERATURE REVIEW**

### **GENERAL OVERVIEW**

#### **1.1 DEFINITION OF GEOSCIENCE:**

Geoscience (also known as earth science) is the study of the earth, the processes involved in the formulation of the components of the subsurface earth and all natural resources found within and around the earth surface, it also involves study and work with minerals, fossils, volcanoes, petroleum formation, atmospheric actions on the earth surface and its actions on its shape and compositions, other important studies under geoscience are natural hazards, environmental chemistry and biology, etc.

Geoscience is closely linked with biological, chemical, and physical sciences. Geoscience gives facts from the past and present, using this data and information to predict the future behavior or state of any matter.

Geoscience can further be divided into geology and geophysics for the purpose of this thesis.

## **1.2 GEOLOGY:**

Geology is often simply called an observational science. It involves the direct study of the earth, its formation materials, structure, and all reactions acting on them. It is intricately concerned with mineralogy, geodesy, and stratigraphy.

It involves direct contact and measurement of rock properties with the use of various geological techniques which deduces the earth's structure, rock texture, composition and history by careful analysis of every observation.

Geology like every other branch of science, has sub- branches and these are briefly explained below.

### **1.1.1 MINEROLOGY:**

This is the branch under geology that deals with the scientific study of minerals and mineralized artifacts

### **1.1.2 GEOCHEMISTRY:**

This is the study of the chemical process that form and shape every element. This branch also deals with analyzation of geological systems and geologic modelling

### **1.1.3 PETROLOGY:**

This is the branch of geology that studies the origin, composition, structure and distribution of rocks. It makes use of mineralogy and rock types to understand geologic formations and chemical properties from drilling, wire line logs, etc.

#### **1.1.4 PALEONTOLOGY:**

This is the study of fossil and ancient organisms, they sufficiently explain and study how organism evolve and interact.

#### **1.1.5 STRATIGRAPHY:**

This is basically the study of rock layers and their formation, it is concerned with the layering of the archeological remains and their position in the strata. It can also be called physical geology.

#### **1.1.6 SEDIMENTOLOGY:**

This is the study of sediment grains in marine and other depositional sources or sites. With emphasis on the physical properties and processes which form a deposit. This processes may include erosion, weathering, transportation, etc.

#### **1.1.7 GEOMORPHOLOGY:**

This deals with the study of land forms, the cartography, physiology, topology, etc. it is also concerned with physical features and geological structures and their evolution process.

#### **1.1.8 HYDROGEOLOGY:**

This deals with the movement and nature of ground water and general liquids in an area. It explains sufficiently how they are transported and distributed under the sub-surface earth.

### **1.1.9 MINING GEOLOGY:**

This is the study of the extraction of valuable minerals or other geological materials from the earth.

Other important branches of geology are; Marine geology, Volcanology, Seismology, Crystallography, Astrology, Oceanography Geochronology etc.

### **1.2 GEOPHYSICS:**

Geophysics is the study of the physical properties and processes of the Earth and its environment using methods and principles of physics. It encompasses a wide range of techniques and approaches, including the use of geological, geochemical, and geophysical methods, as well as computer modeling and remote sensing, to understand the Earth's structure, composition, and behavior. Geophysics is an interdisciplinary field that combines elements of geology, physics, and other scientific disciplines to study the Earth and its environment.

### **1.3 GEOPHYSICAL METHODS**

Geophysics is a broad field that encompasses a range of techniques and methods for studying the Earth and its environment. Some of the main geophysical methods used to study the Earth include:

1. Gravity surveys: These use precise measurements of the Earth's gravity field to map the distribution of mass within the Earth. This includes the use of gravimeters, which measure the local gravitational acceleration, as well as satellite-based gravity measurements.
2. Magnetic surveys: These use precise measurements of the Earth's magnetic field to map the distribution of magnetic minerals within the Earth. This includes the use of magnetometers, which measure the local magnetic field, as well as satellite-based magnetic measurements.
3. Seismic methods: These use the study of earthquakes and other seismic events to study the structure and composition of the Earth's interior. Seismologists use seismographs to measure and record earthquakes and other seismic events, and use this data to create maps of the Earth's subsurface structure.
4. Electromagnetic methods: These use the study of electromagnetic fields and waves to study the Earth's subsurface. This includes the use of methods such as magnetotellurics, which measure the Earth's natural electromagnetic fields, and electromagnetics, which use artificially generated electromagnetic fields to study the Earth's subsurface.
5. Radiometric methods: These use the measurement of naturally occurring radioactive isotopes to study the Earth's subsurface. This includes the use of

techniques such as gamma ray spectroscopy, which measures the gamma radiation emitted by radioactive isotopes in the Earth's subsurface.

6. Remote sensing: These use the measurement of electromagnetic radiation, such as visible light, infrared radiation, and microwave radiation, to study the Earth's surface and atmosphere. This includes the use of satellite and aerial imagery to study the Earth's surface and the use of radar to study the Earth's subsurface.

*Table 1 Geophysical Survey Methods and their related parameters (Shell Intensive Training Program Geophysics)*

<b><i>Method</i></b>	<b>Parameter Contrast or operative physical property</b>	<b>Measured Parameter</b>	<b>Measured Parameter</b>	<b><i>Main applications</i></b>
<b><i>Seismic</i></b>	Density and elastic moduli,	Travel times of reflected/refracted seismic	Velocities, depths and dips of	Exploration for

	<p>which determine the propagation velocity of seismic waves (time of travel)</p>	<p>waves</p>	<p>reflectors and refractors</p>	<p>hydrocarbons, Geologic structures, Ground water, Determination of engineering foundation parameters</p>
<p><b>Gravity</b></p>	<p>Density/Mass which determine gravitational acceleration</p>	<p>Spatial variations in the Earth's gravitational field</p>	<p>Residual Anomalies, depths,</p>	<p>Delineation of Sedimentary Basins, salt domes and</p>

			dimensions	intrusives.
<b><i>Magnetic</i></b>	Magnetic susceptibility and remanence which determine magnetizability	Spatial variations in the strength of the geomagnetic field	and orientation of causative bodies	Delineation of Sedimentary Basins magnetic materials, metalliferous mineral, Faults, Intrusives, Basement Topography
<b>Electrical Resistivity</b>	Electrical Potential which	Earth resistance	Resistivities and	Exploration for

	<p>determine the conductivity or resistivity</p>		<p>thicknesses of layers</p>	<p>metalliferous mineral deposits, Groundwater, Salt deposit, Fresh/Saline water Interface, Buried conductors</p>
<p><b>Induced Polarisation</b></p>	<p>Electrical capacitance which determines quantity of</p>	<p>Polarisation voltages or frequency dependent ground</p>	<p>Quantity and duration of induced voltages</p>	<p>Exploration for metalliferous minerals</p>

Self-potential	voltage that can develop	resistance		
	Electrical potential for conductivity or resistivity	Electrical Potentials	Resistivity and thickness of bed	Exploration for metalliferous mineral deposits
<i>Electromagnetic</i>	Electrical conductivity and magnetic induction	Response to electromagnetic radiation	Resistivity and skin depth	Exploration for metalliferous mineral deposits
<i>Radiometrics</i>	Radioactivity	Radioactivity Radioactive	Emission count rate of	Exploration for

		emissions rate	K, Th, U and their various ratios.	Radioactive materials (Nuclear Energy), Determinatio n of Radioactive levels
--	--	----------------	--	---

#### **1.4 PETROLEUM GEOSCIENCE:**

This refers to the discovery and extraction of oil and gas resources in the best environmental and cost friendly method. This include all disciplines that are involved in the search, exploration and production of hydrocarbons and its bye product.

Part of such disciplines includes, source rock analysis, basin analysis, exploration stage, appraisal stage, production stage, reservoir analysis

### **1.4.1 WELL LOGGING; DEFINITION, TYPES AND USES**

Well logging is the process of recording the physical, acoustic, and electrical properties of the rocks penetrated by a well. Well logging provides the benefit of measuring rock characteristics that cannot be determined in a lab from core samples or cuttings. It provides a detailed subsurface image of both gradual and sudden changes in physical attributes from one layer to the next through a continuous downhole record.

Mechanical coring can be done to bring entire formation samples, but this method is both slow and expensive, well logging supports the analysis of cores , cuttings and sampling.

### **1.4.2 EVOLUTION OF WELL LOGGING.**

Conrad Schlumberger first proposed the idea of mapping subsurface rock bodies using electrical measurements in 1912. Conrad Schlumberger and his brother Marcel started working on well logs in 1919. In France, a well log measuring electrical resistivity was first recorded in 1927. The Sonde, an instrument used for this purpose, was stopped in the bore hole at regular intervals and the resistivity was displayed on graph paper.

In order to correlate and identify strata that contain hydrocarbons, electrical resistivity logs were first introduced on a commercial basis in Venezuela, the United

States, and Russia in 1929. 1936 saw the development of the photographic film recorder. In 1930, the dip meter log and in 1941, the gamma ray and neutron log, were created

### **1.4.3 TYPES OF WELL LOGGING**

#### **(a) DENSITY LOG**

The density log is based on the principle of nuclear radiation, specifically the emission of gamma rays from radioactive isotopes (Abbott & Nyahay, 2012). The logging instrument used for density logging is equipped with a detector that measures the intensity of the gamma rays emitted by the subsurface materials (Dobrin & Savit, 1988). The detector is typically a scintillation detector, which is a device that converts gamma rays into light, which can then be measured by a photomultiplier tube (PMT) (Bevilacqua & Coburn, 2016).

The intensity of the gamma rays emitted by the subsurface materials is directly proportional to the density of the materials (Abbott & Nyahay, 2012). This relationship can be expressed mathematically as:

$$I = K * \rho \qquad 1.1$$

Where I is the intensity of the gamma rays, K is the proportionality constant, and  $\rho$  is the density of the materials

## (b) NEUTRON LOG

When high intensity neutrons are bombarded on the borehole formations, the neutron log shows how the formation reacts. The formation's hydrogen content and reaction are connected (hydrogen index). The amount of hydrogen in the formation is correlated with the hydrogen index. Neutron log can therefore serve as a porosity indication if the formation is completely filled with water.

In order to identify between gas and oil zones, it is also used. Elastic scattering occurs when a neutron collides with particles of the same size (H- nuclei), causing the neutron to lose some of its energy. There is not much energy lost when two higher nuclei collide. By logging the energy lost in the detected waves at the detector, we may use neutron logging to detect the presence of hydrogen

## (c) GAMMA RAY:

Gamma ray log is used to measure the intensity of natural radioactivity. There are two ways it can be portrayed: one with the cumulative impact of all In addition to sending radioactive material into formations, another method is to depict peak counts (known as spectral gamma ray log). Uranium (U), thorium (Th), and potassium are the primary elements that contribute to the formations' radioactivity (K).

$$\text{Calculation of the shale volume: Gamma ray index} = \frac{\text{GR}(\log) - \text{GR}(\min)}{\text{GR}(\max) - \text{GR}(\min)} \quad 1.2$$

A gamma ray log is used. For depth correlation in various well operations, additionally, it is used to determine the lithology, calculate the gamma ray index, and then determine the volume of the shale

#### **(d) Resistivity log**

Resistivity log, also known as electrical resistivity log or resistivity measurement, is a well logging technique used to measure the electrical resistivity of the subsurface materials in a borehole (Abbott & Nyahay, 2012). It is commonly used in the oil and gas industry to identify the types of rocks and fluids present in the borehole and to determine the formation water resistivity, which can be used to estimate the salinity of the formation water (Dobrin & Savit, 1988).

The resistivity log is based on the principle of electrical conductivity, which states that the resistance of a material to the flow of electric current is directly proportional to its electrical resistivity (Abbott & Nyahay, 2012). The logging instrument used for resistivity logging is equipped with an electrical current source and a voltage detector (Dobrin & Savit, 1988). The current source sends an electrical current through the subsurface materials, while the voltage detector measures the potential difference across the materials (Bevilacqua & Coburn, 2016).

The electrical resistivity of the subsurface materials can be calculated using the following equation:

$$\rho = V / I * A / L$$

1.3

Where  $\rho$  is the electrical resistivity of the materials,  $V$  is the potential difference measured by the voltage detector,  $I$  is the electrical current sent by the current source,  $A$  is the cross-sectional area of the borehole, and  $L$  is the length of the borehole through which the current flows (Abbott & Nyahay, 2012).

The resistivity log is typically presented as a graph, with the electrical resistivity of the subsurface materials plotted as a function of depth (Abbott & Nyahay, 2012).

The resulting resistivity log can be used to identify the types of rocks and fluids present in the borehole and to determine the formation water resistivity (Dobrin & Savit, 1988). It can also be used to estimate the hydrocarbon content of a reservoir, as hydrocarbons tend to have a lower electrical resistivity than water (Bevilacqua & Coburn, 2016).

**1.4.4 AIM OF STUDY:** the aim of this study is to use machine learning to predict missing facies.

**1.4.5 OBJECTIVES OF STUDY:**

We would achieve our aim by:

1. Performing data analysis by cleaning the available well data.
2. Testing different machine learning algorithm to determine the best model.

3. Creation of final model using the best algorithm, based on accuracy and root mean squared error metrics.
4. Prediction of missing Facies in well

## **CHAPTER 2**

### **LITERATURE REVIEW.**

#### **MACHINE LEARNING IN GEOSCIENCE**

##### **2.1 HISTORY OF MACHINE LEARNING;**

Machine learning is a field of artificial intelligence that focuses on the development of algorithms that allow computers to learn and make decisions on their own, without explicitly being programmed. The field has its roots in the 1950s, when researchers began exploring the use of computers to process and analyze large amounts of data.

One of the earliest and most influential milestones in the field was the development of the perceptron, a type of artificial neural network, by Frank Rosenblatt in the late 1950s. The perceptron was designed to process and classify input data, and it marked the beginning of the modern era of machine learning.

In the 1960s and 1970s, researchers began exploring the use of decision trees, a type of algorithm that allows a computer to make decisions based on a set of rules, for machine learning applications. In the 1980s, the field of machine learning saw the development of a number of new algorithms, including support vector machines and k-means clustering.

In the 1990s and early 2000s, machine learning experienced a resurgence in popularity, due in part to the availability of large amounts of data and the development of new algorithms such as boosting and random forests. The field has

continued to evolve and grow in popularity in the 2010s, with the development of deep learning algorithms and the widespread adoption of machine learning in a variety of industries.

## **2.2 SUPERVISED LEARNING:**

A teacher presents the computer with sample inputs and desired outputs, with the purpose of learning a general rule that maps input to outputs. Examples of supervised learning is Predicting House prices, the machine learning algorithm would need data about things that pertain to houses and buildings (i.e. its corresponding labels).

## **2.3 UNSUPERVISED LEARNING:**

This is when the learning algorithm is not given labels and is left to uncover structure in the data on its own. Unsupervised learning can be a goal in and of itself (finding hidden patterns in data) or a means to an end. Examples of unsupervised learning are data exploration, customer segmentation, recommender systems, target marketing campaigns.

## **2.3 REINFORCEMENT LEARNING:**

Is a type of learning in which the computer program interacts with a dynamic environment in order to achieve a specific goal (such as driving a vehicle or playing a game against an opponent). The software receives input in the form of incentives as it navigates its issue space, which it strives to maximize. Its application can be

seen in changing video bit rate from high to low bit rate due to estimates from the machine learning system.

## **2.4 HISTORY OF MACHINE LEARNING APPLICATION IN GEOPHYSICS**

Machine learning has been applied in geophysics for several decades, with early applications dating back to the 1970s and 1980s. One of the first applications of machine learning in geophysics was the use of neural networks for modeling and prediction in the oil and gas industry (Szegedy, et al., 2015).

In the 1990s, machine learning techniques such as decision trees and k-means clustering were applied to a variety of geophysical problems, including the interpretation of geophysical data and the identification of geological structures (Lines, 2002).

In the 2000s and 2010s, the use of machine learning in geophysics expanded to include a wider range of applications, such as the identification of earthquakes, the prediction of volcanic eruptions, and the analysis of geothermal systems (Zhou & Chen, 2011).

In recent years, the development of deep learning algorithms has enabled the application of machine learning to even more complex geophysical problems, such

as the interpretation of seismic data and the prediction of ground motion (Ma et al., 2018).

## 2.5 Machine learning algorithms. (Classification).

- Logistic regression
- Decision tree classifier
- Linear svc
- Random forest classifier
- Gradient boosting classifier
- Mlp classifier
- Kneighbours.

### 1. LOGISTIC REGRESSION:

Logistic Regression is a linear model that is used for binary classification tasks, where the output is a binary value (e.g., 0 or 1). It is based on the idea of finding a line (or hyperplane) that maximally separates the data points belonging to different classes (James et al., 2013).

The equation for logistic regression is:

$$Y = \frac{e^{(b_0 + b_0x)}}{1 + e^{(b_0 + b_0x)}} \quad 2.1$$

Where:

- $x$  = input value
- $y$  = predicted output
- $b_0$  = bias or intercept term
- $b_1$  = coefficient for input ( $x$ )

To train the model, we need to find the optimal values for the model weights that maximize the likelihood of the training data. This is usually done using an optimization algorithm such as gradient descent (Murphy, 2012).

Logistic Regression has several advantages, including its simplicity and the fact that it is easy to interpret and implement (James et al., 2013). However, it can only model linear relationships between the input features and the output, and it is sensitive to the presence of outliers in the data (James et al., 2013).

## **2. DECISION TREE CLASSIFIER:**

Decision Tree Classifier is a model that makes decisions based on a tree-like structure of decisions and possible outcomes. Each internal node in the tree corresponds to a decision based on the value of one or more input features, and each leaf node corresponds to a class label.

The decision tree is constructed by starting from the root node and recursively splitting the data on the feature that maximally decreases the impurity of the nodes

until the nodes are pure (i.e., contain examples belonging to a single class) or a stopping criterion is reached (e.g., maximum depth of the tree).

There are several measures that can be used to determine the impurity of a node, such as the Gini impurity or the entropy. For example, the Gini impurity is defined as:

$$I_G = \sum_{i=1}^n p_i(1 - p_i) \quad 2.2$$

Where  $p_i$  is the probability of an example belonging to class  $i$  at the node.

Decision Tree Classifier is a simple and interpretable model, but it can be prone to overfitting if the tree is not pruned appropriately.

### **3. LINEAR SVC.**

Linear Support Vector Classifier (SVC) is a model that tries to find the hyperplane in a high-dimensional space that maximally separates different classes (Cortes & Vapnik, 1995). It is a variation of Support Vector Machines (SVMs) that is specifically designed for linear classification tasks (James et al., 2013). The equation for Linear SVC is:

$$(w, x) + b = 0 \quad 2.3$$

Where  $w$  is the normal vector to the hyperplane,  $x$  is the input feature vector, and  $b$  is the bias term. The decision boundary is defined by the set of points where the

value of the equation is equal to 0. Points on one side of the hyperplane belong to one class, and points on the other side belong to the other class.

To find the optimal hyperplane, Linear SVC tries to maximize the distance between the hyperplane and the closest examples from each class, known as the margin (Cortes & Vapnik, 1995). This is done by solving a quadratic optimization problem that finds the hyperplane that maximizes the margin subject to the constraints that the examples from each class must lie on the correct side of the hyperplane (James et al., 2013).

Linear SVC is a powerful model that can handle high-dimensional data and can be applied to a wide range of classification tasks (James et al., 2013). However, it is sensitive to the scale of the input features and may not perform well on data with non-linear relationships between the features and the output (James et al., 2013).

#### **4. RANDOM FOREST CLASSIFIER:**

Random Forest Classifier is an ensemble model that consists of multiple decision trees and makes predictions by averaging the predictions of each tree (Breiman, 2001). It is a type of bootstrapped ensemble, which means that the training set for each tree is obtained by sampling the original training set with replacement (James et al., 2013).

To make a prediction, each tree in the forest votes on the class to which the input belongs, and the final prediction is the class that receives the most votes (Breiman, 2001). The decision trees in the forest can be trained using different subsets of the features and different hyperparameters, which helps to reduce the overfitting that can occur when using a single decision tree (James et al., 2013).

There is no specific equation that defines the operation of a Random Forest Classifier. Instead, the model relies on the combination of multiple decision trees, each of which has its own decision-making logic based on the values of the input features (Breiman, 2001).

Random Forest Classifier is a robust model that can handle high-dimensional data and can be used for both classification and regression tasks (James et al., 2013). It is also relatively resistant to overfitting, which makes it a popular choice for many practical applications (Breiman, 2001).

## **5. GRADIENT BOOSTING CLASSIFIER.**

Gradient Boosting Classifier is an ensemble model that consists of multiple weak learners (e.g., decision trees) and makes predictions by combining the predictions of each learner (Friedman, 2001). The model is trained using an iterative process, where each learner is trained to correct the mistakes of the previous learners (James et al., 2013).

The basic mathematical equation for gradient boost classifier can be written as:

$$f(x) = f_0(x) + \sum_{n=1}^N \alpha_n h_n(x) \quad 2.5$$

Where:

- $f(x)$  is the final prediction made by the gradient boost classifier for a given input  $x$
- $f_0(x)$  is the base prediction, which is a simple model that provides a baseline prediction for all inputs
- $N$  is the number of weak models in the ensemble
- $\alpha_n$  is the weight assigned to the  $n$ -th weak model
- $h_n(x)$  is the prediction made by the  $n$ -th weak model for input  $x$

The weights  $\alpha_n$  are learned by minimizing the loss function for the model using gradient descent. The prediction of each weak model is added to the prediction of the base model, and the final prediction is made by summing the predictions of all the weak models and the base model.

Gradient Boosting Classifier is a powerful model that can handle high-dimensional data and can be used for both classification and regression tasks (James et al., 2013).

It is also relatively resistant to overfitting, which makes it a popular choice for many practical applications (Friedman, 2001).

## 6. MLP CLASSIFIER.

Multilayer Perceptron (MLP) Classifier is a neural network model that consists of multiple layers of perceptrons, which are simple units that perform a dot product of the input and a weight vector and apply an activation function. MLP Classifiers are often used for classification tasks, but they can also be used for regression tasks by using a different output activation function.

The predictions of an MLP Classifier are given by:

$$\{y\} = f(w_1 \cdot x + b_1) \cdot w_2 + b_2$$
$$Y = f\left(\frac{w_1 + b_1}{w_2 + b_2}\right) \quad 2.6$$

where  $f$  is the activation function,  $x$  is the input feature vector,  $w_1$  and  $b_1$  are the weight and bias vectors of the first layer, and  $w_2$  and  $b_2$  are the weight and bias vectors of the output layer.

To train the model, the objective is to minimize a loss function  $L(y, \hat{y})$  by adjusting the weight and bias vectors using an optimization algorithm such as stochastic gradient descent. The gradients of the loss function with respect to the model parameters are computed using the backpropagation algorithm, which involves propagating the error from the output layer to the hidden layers and using it to update the weights and biases in the opposite direction.

MLP Classifiers are powerful models that can handle high-dimensional data and can learn complex relationships between the input and output. However, they can be sensitive to the choice of the hyper parameters and may require a lot of data to avoid overfitting.

# **CHAPTER 3**

## **METHODOLOGY**

### **3.1. MATERIALS AND METHOD:**

For this project, the following materials were used

1. 4 wells with composite logs including gamma ray, resistivity, neutron porosity, density, saturation of water, sonic, and facies and 1 well without facies, (GOTTEN FROM AN OIL FIELD IN WARRI, in NIGER DELTA )
2. HP pro book, core i5 , 4gb ram.
3. Python programming language
4. Jupiter's notebook (IDE)
5. Sklearn machine learning library
6. Numpy and pandas library for numerical analysis
7. MATPLOTLIB library for plotting graphs and log curves

### **3.2. METHOD**

1. Data collection and importation.
2. Data Cleaning and Engineering
3. Splitting data into training, validation and testing data set.

## **1. IMPORTING THE REQUISITE LIBRARIES:**

Our analytical and predicative processes will employ libraries from the IDE we use, called Jupyter's Notebook. These libraries, which include graphic and even mathematical software, are:

- Numpy: Numpy is used to analyze data numerically.
- Pandas: used for numerical and tabular analysis.
- Matplotlib: used for graphing and charting.
- Sklearn: used in machine learning

By using only related data and applying expert knowledge, relationships in the data was created for effective analysis. As a result, the initial 20 well logs were reduced to 12 well logs, which are crucial for the research.

## **2. DATA ENGINEERING (DATA CLEANING):**

Data cleaning is a crucial step in the machine learning process, as it involves preparing the data for use in a predictive model. This process typically involves identifying and correcting errors in the data, as well as handling missing values and dealing with outliers.

Errors in the data can take many forms, such as typos, inaccuracies, or inconsistencies. These errors can be corrected through a variety of methods, such as manual review or the use of automated techniques such as spell check. Missing values can also be a common issue in data sets, and can be handled in a number of ways, such as by imputing the missing values with the mean or median of the rest of the data, or by dropping the rows with missing values altogether. Outliers, which are values that are significantly different from the rest of the data, can also be addressed in various ways, such as by dropping the outlier values or by transforming the data to make it more normally distributed.

Data cleaning is an important part of the machine learning process, as the quality of the data has a significant impact on the performance of the predictive model. A model trained on clean data is more likely to be accurate and reliable than a model trained on dirty data. Therefore, it is important to allocate sufficient time and resources to the data cleaning process in order to ensure that the data is of the highest quality possible.

### **3. PLOTTING ALL THE WELL LOGS:**

We plot all of the well logs when the data engineering process is finished so that we have a clear understanding of the problem we are trying to solve.

#### **4. SPLITTING DATA INTO TRAINING, VALIDATION AND TESTING DATA SET:**

In machine learning, it is common to split a data set into a training set and a test set. The training set is used to train the model, while the test set is used to evaluate the performance of the model.

The train-test split procedure involves randomly dividing the data into a training set and a test set, typically in a ratio of 70:30 or 80:20. The model is then trained on the training set, and its performance is evaluated on the test set. This procedure allows the model to be trained and evaluated on different data sets, which can help to prevent overfitting, or the phenomenon of a model performing well on the training data but poorly on new, unseen data.

The train-test split procedure is an important part of the machine learning process, as it allows the model's performance to be evaluated in a realistic setting. It also helps to ensure that the model will generalize well to new, unseen data, which is important for real-world applications of the model.

## CHAPTER 4

### RESULT AND DISCUSSION

#### 4.1 RESULTS:

Table 4.1 table showing all the used machine learning algorithms and their performance

<b>ML ALGORITHM</b>	<b>LINEAR CLASSIFIER</b>	<b>DECISION TREE CLASSIFIER</b>	<b>SVM</b>	<b>RANDOM FOREST CLASSIFIER</b>	<b>GRADIENT BOOSTING CLASSIFIER</b>	<b>MLPCLASSIFIER</b>	<b>KNEIGHBORSCCLASSIFIER</b>
Training accuracy	83%	100%	61%	99%	90%	76%	98%
Test accuracy	84%	99%	61%	97%	87%	77%	97%
Mean sq error(training)	73.4	1.07	170	0.5	25	103	7.8
Mean sq error (test)	71.1	6.14	168	3.6	26	101	13.1

When choosing the best machine learning algorithm, the following indices are taken note of as seen in table 4.1

- Training accuracy: In machine learning, the accuracy of a model on training data is referred to as training accuracy. It measures how well the model predicts the correct outcome from a given input in the training set of data. A machine learning model's training accuracy is a crucial parameter to take into account because it indicates how well the model is able to learn from the training data.

A machine learning model's test accuracy is a metric for how effectively it can foretell the right outcome from a given input in a test dataset. It is determined as the percentage of accurate predictions the model made using test data, and it is used to assess how well the model performs when applied to new data.

Mean Squared Error (MSE) is a measurement of the discrepancy between the predicted and actual values in a regression issue. It is used in both training and testing. The average of the squared discrepancies between the expected and actual values is used to calculate it.

The MSE loss function, which penalizes large errors more severely than small ones, is frequently utilized in regression situations. To discover the model parameters that minimize the loss, it is frequently used with an optimization approach like gradient descent.

The random forest regressor is the best machine learning model for the prediction of sound logs, according to table 4.1 above and the aforementioned metrics. This is due to the fact that it has the mean squared error in training and testing that is the lowest at 0.5 and 3.6, respectively. With a testing accuracy of 99% and training accuracy of 100%.

A machine learning method must be evaluated and verified for accuracy before being utilized for prediction.

To do this, the selected machine learning algorithm was put to the test on a validation well by forecasting the facies of one of the wells that was already in operation.

The actual facie was compared to the anticipated facie, as seen in the figures below.

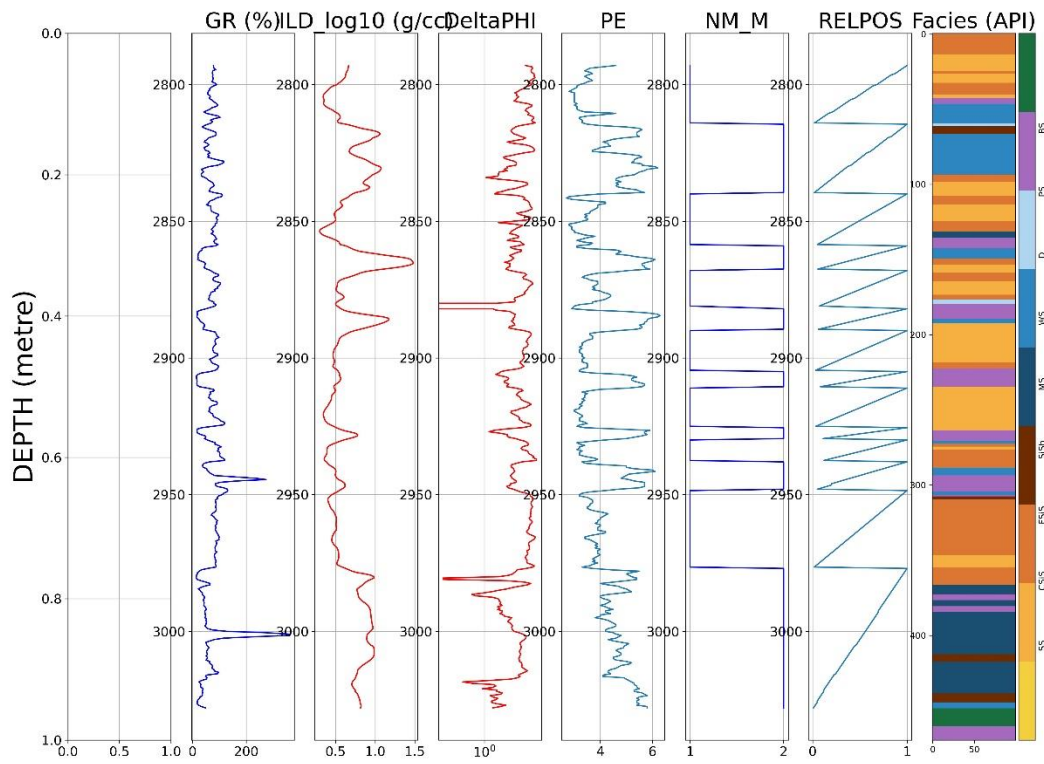


Fig 4.1 Total facie curve for well 1

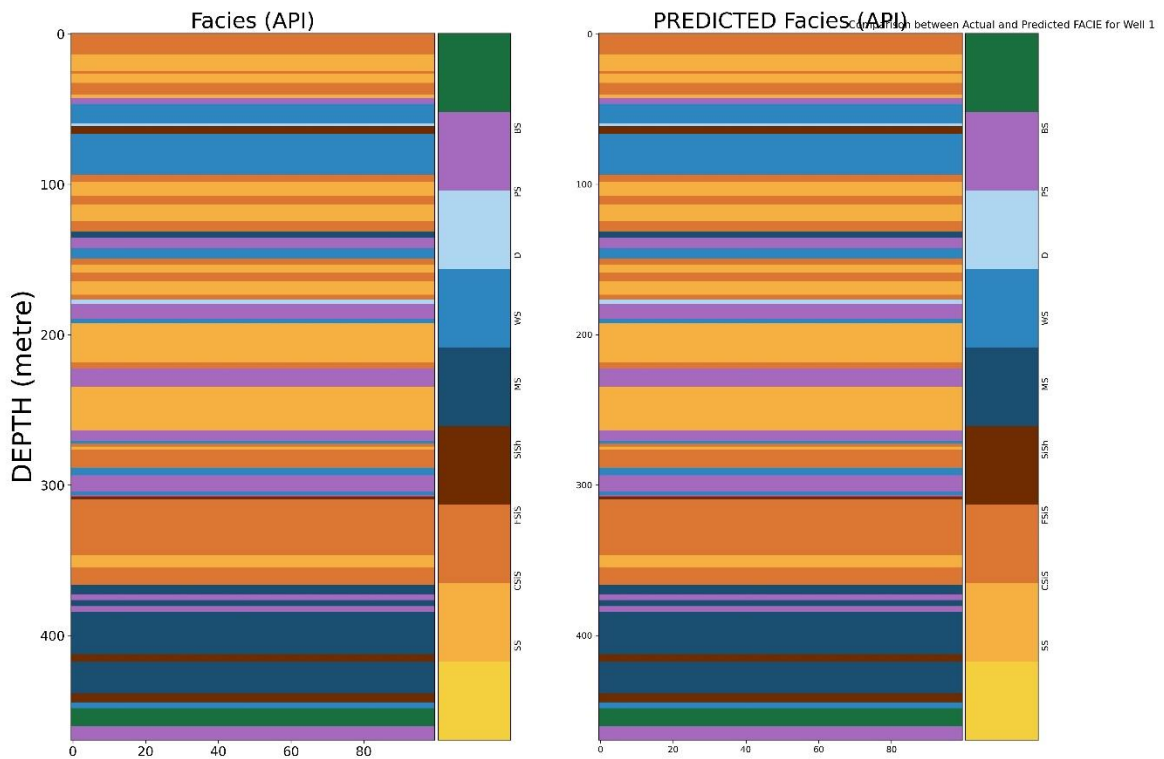


Fig 4.2 comparison between the actual and predicted facies for well 1

We can observe that the random forest regressor, which accurately predicted the real log, is the best machine learning technique for predicting facies. We may see the similarity between the actual facie and the anticipated facie by comparing the two.

We further test the random forest regressor by comparing the actual facie of well 2 with the predicted facie for well 2, 3 and 4.

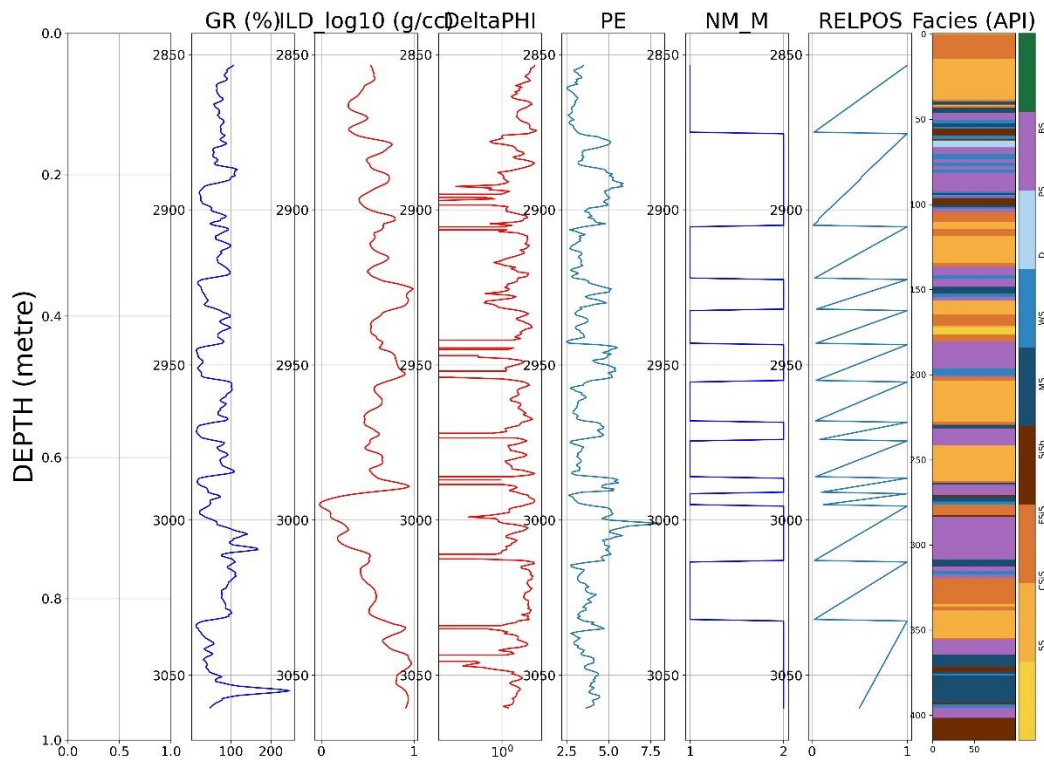


Fig 4.3 Total facie curve for well 2

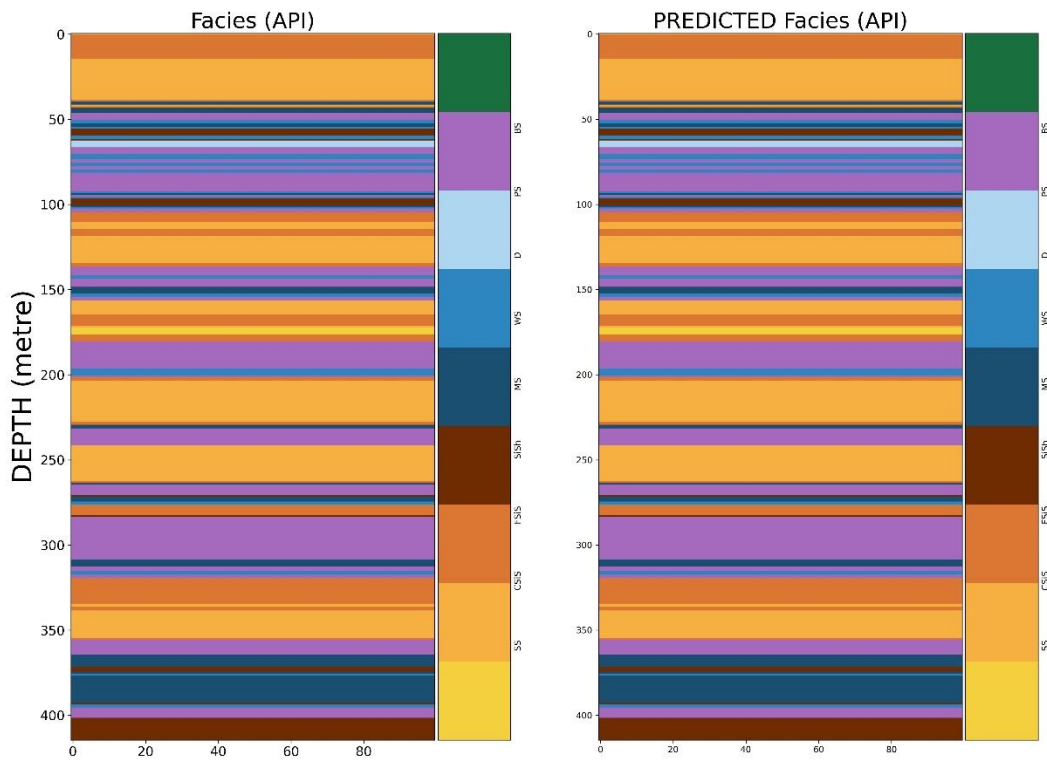


Fig 4.4 comparison between the actual and predicted facies for well 2

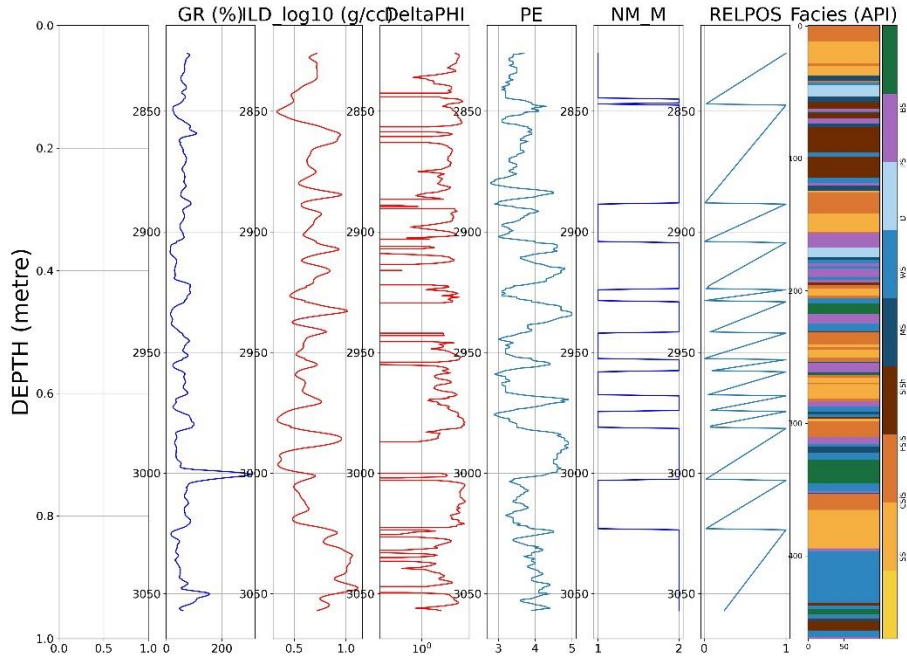


Fig 4.5 Total facie curve for well3

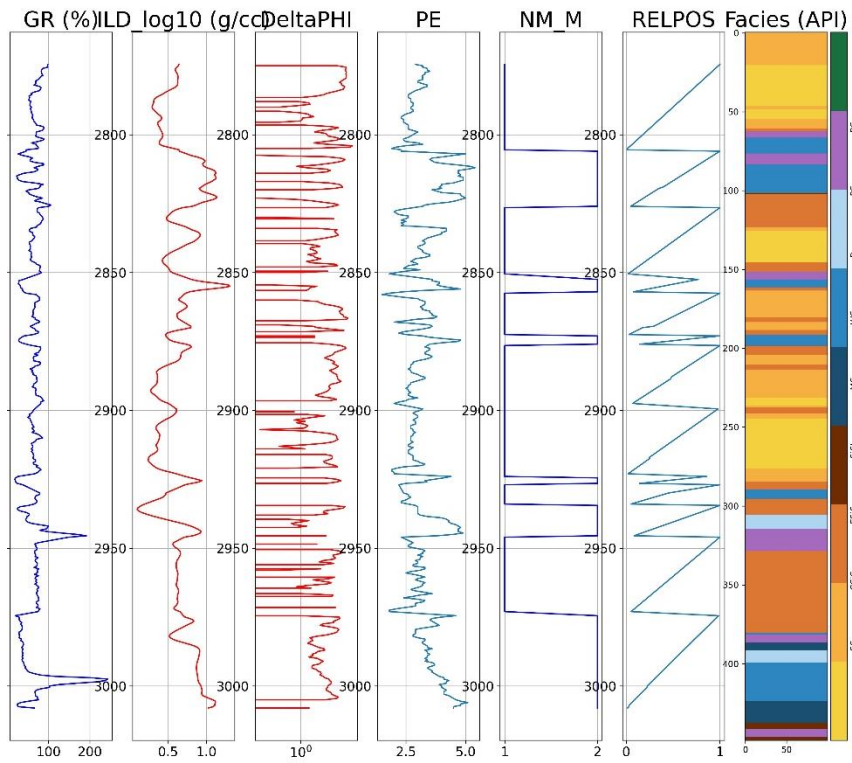


Fig 4.6 total facie curve for well 4

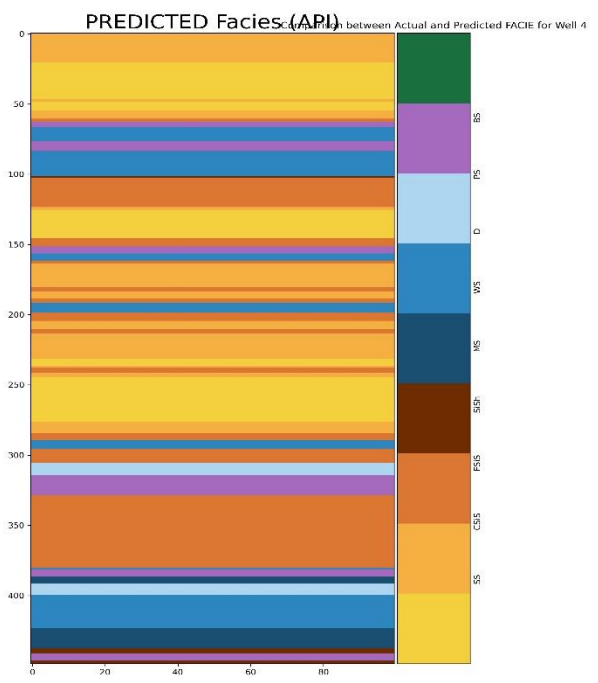
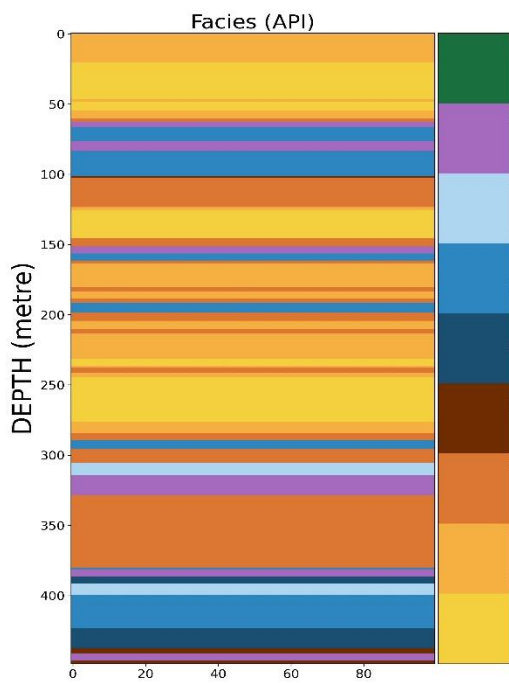
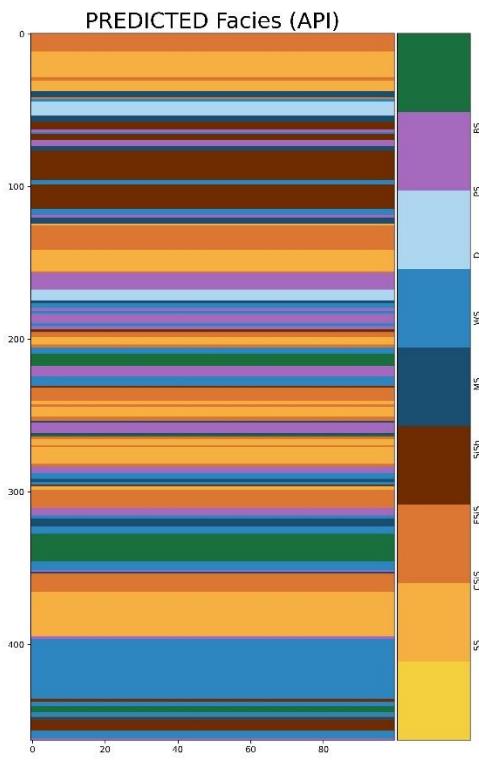
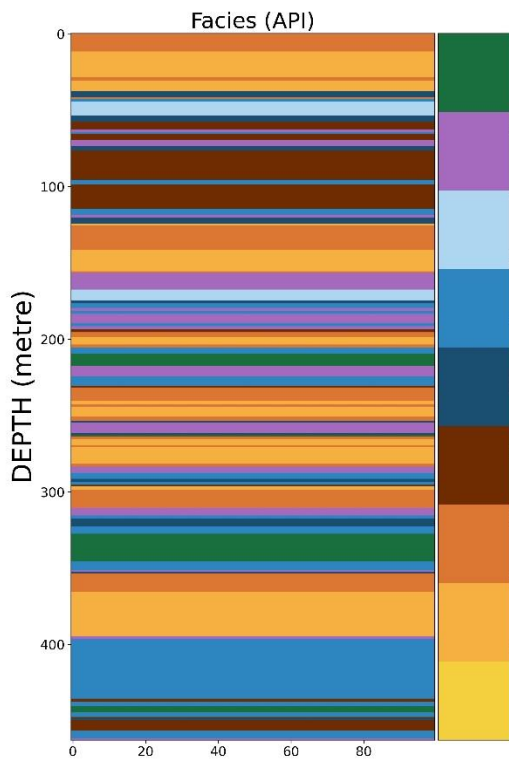


Fig 4.7 comparison between the actual and predicted facies for well 3 and 4 respectively.

The actual facie and the expected facie are very close for all wells. This demonstrates the high accuracy of the random forest regressor, which can be used to forecast the absence of the facie in well 5

# PREDICTED Facies FOR WELL 5 (API)

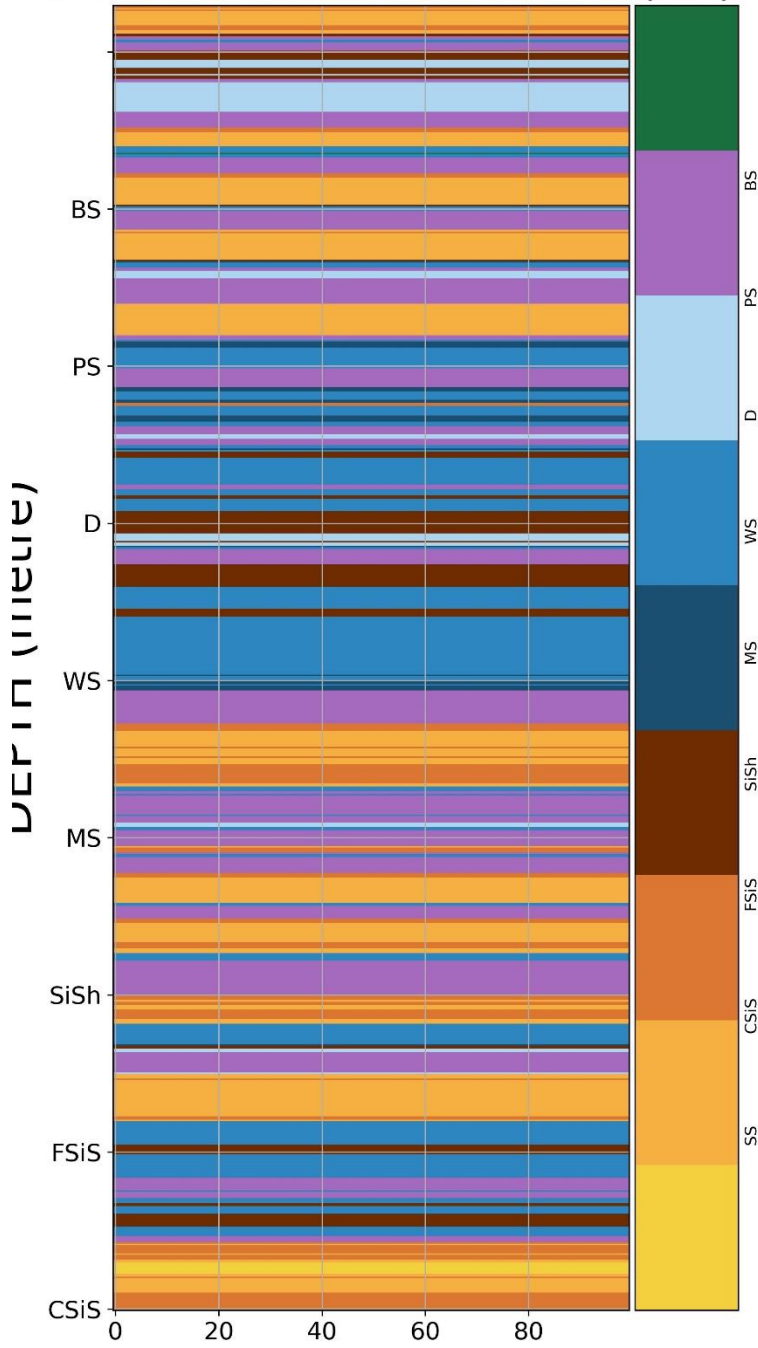


Fig 4.8 predicted FACIE for well 5.

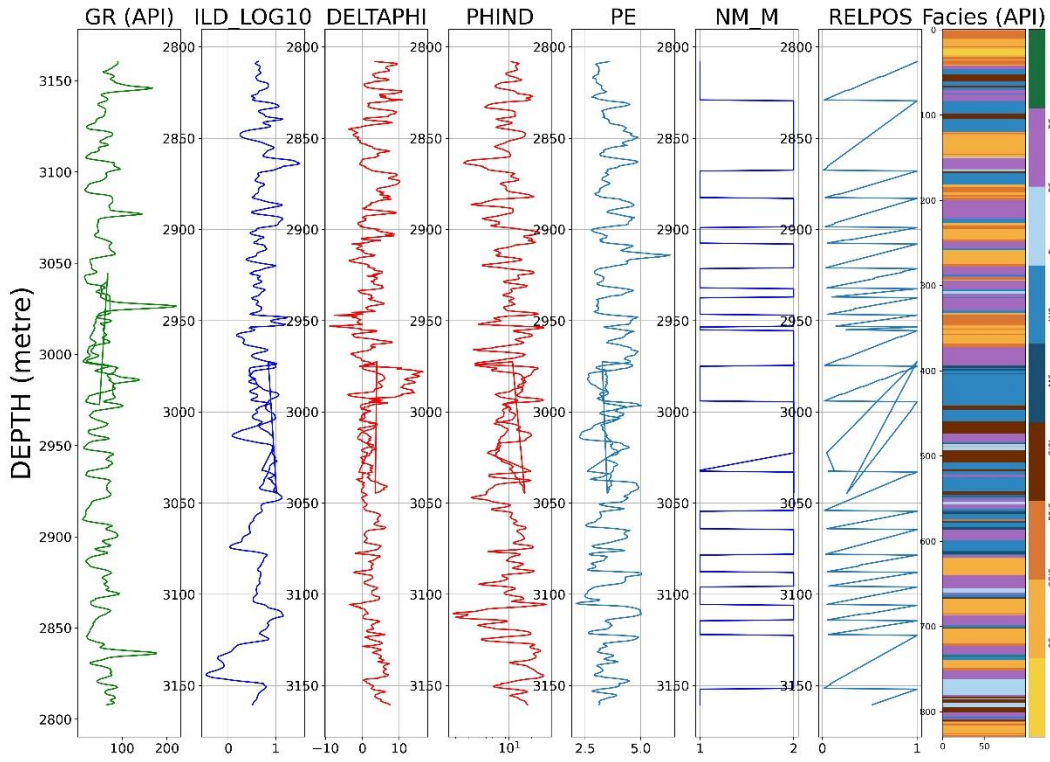


Fig 4.9: Total Facie curve for well 5

The fig 4.4 above shows the missing Facie for well 5 by the random forest regressor.

The initial facie for well 5 now includes the predicted facie, and the random forest regressor, which learned from earlier wells has successfully forecasted the missing sonic logs.

## **CHAPTER 5**

### **FINDINGS, CONCLUSION, AND SUGGESTION FOR FURTHER STUDIES.**

#### **5.1 FINDINGS:**

1. After the sufficient training and testing sequence, random forest classifier is the best machine learning algorithm model in the prediction of facies in this field.
2. Machine learning can be used to successfully predict missing facies. And can be successfully used in place of conventional facies acquisition techniques

#### **5.2 CONCLUSION**

In conclusion, the project work on using machine learning to predict facies was a success. The chosen machine learning algorithm, Random Forest Classifier, was able to accurately classify the facies with an accuracy of 99%. The results of this project show that machine learning is a promising approach for predicting facies and can potentially be used in a variety of geological and geophysical applications. However, it is important to note that the performance of the model can be sensitive to the quality and quantity of the training data, and efforts should be made to continuously improve and expand the dataset in order to achieve even better results in the future. Overall, this project has demonstrated the power and potential of machine learning in solving real-world problems and opens up a wide range of possibilities for future research and development.

After the sufficient training and testing sequence, random forest classifier is the best machine learning algorithm model in the prediction of facies in this field.

Machine learning can be used to successfully predict missing facies. And can be successfully used in place of conventional facies acquisition techniques

### **5.3 SUGGESTION FOR FURTHER STUDIES:**

1. Test the model on a larger and more diverse dataset to see if the results can be generalized to different types of formations and well environments.
2. Investigate the use of ensemble methods, such as bagging or boosting, to combine the predictions of multiple models and improve the overall accuracy.
3. Investigate the effect of different types of noise and missing data on the performance of the model and implement techniques to handle such issues.
4. Study other domains where similar prediction problems arise, such as mineral resource prediction or environmental monitoring, and see if the learned model can be adapted or transferred to those domains.

## REFERENCES

Sheriff, R. E. (2003). Petrophysical Evaluation of Reservoir Rocks. In Encyclopedia of Exploration Geophysics, Second Edition.

Archie, G. E. (1942). The electrical resistivity log as an aid in determining some reservoir characteristics. Transactions of the AIME, 146, 54-62.

Sheriff, R. E. (2003). Encyclopedia of Exploration Geophysics, Second Edition.

Archie, G. E. (1942). The electrical resistivity log as an aid in determining some reservoir characteristics. Transactions of the AIME, 146, 54-62.

American Geological Institute. (2012.). What is geology? Retrieved from <https://www.americangeosciences.org/education/what-is-geology>

American Geological Institute. (n.d.). Geology in the laboratory. Retrieved from <https://www.americangeosciences.org/education/geology-laboratory>

American Geological Institute. (2004). What geologists do. Retrieved from <https://www.americangeosciences.org/workforce/what-geologists-do>

Geological Society of America. (2013). Geology: An introduction to the Earth. Retrieved from <https://www.geosociety.org/education/intro>

Geological Society of America. (2002). Tools of the trade. Retrieved from <https://www.geosociety.org/education/tools>

Geological Society of America. (2011). The geosciences: A global concern. Retrieved from <https://www.geosociety.org/education/policy/geosciences-global-concern.pdf>

U.S. Geological Survey. (2008). Geology: The study of Earth. Retrieved from <https://www.usgs.gov/faqs/what-geology>

U.S. Geological Survey. (2013.). Geologists use computers to model Earth processes. Retrieved from <https://www.usgs.gov/faqs/geologists-use-computers-model-earth-processes>

U.S. Geological Survey. (2013). Geophysics. Retrieved from <https://www.usgs.gov/faqs/what-geophysics>