Reservoir Characterization of Zamzama Gas Field by using Machine Learning and Artificial Intelligence

BY

Muhammad Bilal Malik

M.Phil. GEOLOGY 2020-2022

DEPARTMENT OF EARTH SCIENCES QUAID-I-AZAM UNIVERSITY ISLAMABAD, PAKISTAN

CERTIFICATE

This dissertation is submitted by **Muhammad Bilal Malik** S/o Malik Muhammad Akram is accepted in its present form by the Department of Earth Sciences, Quaid-e-Azam University Islamabad as it satisfies the requirement for the award of M.Phil. Degree in Geophysics.

RECOMMENDED BY:

Dr. Matloob Hussain _____

(Supervisor)

Dr. Aamir Ali _____

(Chairman Department of Earth Sciences)

External Examiner _____

Department of Earth Sciences

Quaid-e-Azam University Islamabad, Pakistan

DEDICATED TO

MY PARENTS, BROTHERS AND LOVING, CARING AND SWEET

FRIENDS

AND ALL THOSE WHO HELPED ME IN THIS WORK

ACKNOWLEDGMENT

First and foremost, all praises to Allah Almighty, the most beneficent and the most merciful. Secondly, my humblest gratitude to the Holy Prophet Muhammad (Peace Be Upon Him) whose way of life has been a continuous guidance and knowledge of humanity for me. This thesis appears in its current form due to the assistance and guidance of several people. It gives me great pleasure to express my gratitude to all those who supported me and have contributed in making this thesis possible.

I am also extremely grateful to my honorable supervisor **Dr. Matloob Hussain** for their wide knowledge and logical way of thinking have been of great value for me. Their personal guidance has provided a good basis for this study. I must acknowledge the cooperation of all my teachers whose direction and support have been the source of my success.

I would like to pay my colossal respect to my friend **Sher Afgan**. This work could never have been completed without him. No matter what time it was, he was always there for assistance with warm welcome.

I would like to acknowledge my seniors and friends for their help and suggestions during my work. I do express my sincere thanks to Palwasha Shahzad Rathore, Maha Ali Haider, Muhammad Umais and Anees ur Rehman for their sincere guidance, help, moral support and encouragement

I am thankful to my parents for moral and financial support. I am indebted to my loving parents, who encouraged and motivated me to face the challenges of life during my academic career. I am also highly indebted to my brothers and specially my seniors for their love and cooperation.

Last but not the least, I am thankful to all my friends and fellows who encouraged and motivated me throughout my academics.

Muhammad Bilal Malik

M.Phil. Geology

2020-2022

Abstract

The reservoir characterization is an important step in the exploration of oil and gas. Reservoir characterization enables the exploration scientists to estimate the properties of reservoir. This study is conventionally performed by using the seismic data along with well log data. These datasets provides the inFormation about the subsurface geology and the rock Formations.

The advancement in computational techniques and increasing complexity of the hydrocarbon reservoirs motivated the geoscientists to introduce the advanced computational techniques like artificial intelligence and machine learning in the workflow of hydrocarbon exploration workflow. The machine learning techniques provides the efficient tool for the intelligent analysis of huge datasets of subsurface. It also automates the analysis process that greatly reduces the chances of personal error in seismic and petrophysical interpretation. The present study involves the seismic interpretation using the machine learning techniques. The prediction of missing well curves were also performed by using the available dataset.

Random forest, Support Vector Machine, Decision Tree and Extreme Gradient Boosting (Xgboost) are implemented with Python programming language and reservoir characterization is done with Python programming language 3.9.0.

This study employed a total of five wells, with extensive petrophysical interpretation performed first. The machine learning random forest system was trained on two selected wells in order to forecast the essential petrophysical parameters across the cube, resulting in an 80% match. Facies modelling was conducted out using these petrophysical volumes as input to the K-Mean clustering technique.

Contents

LIST OF FIGURES

List of Tables

Chapter 1

1 INTRODUCTION

The term Reservoir Characterization (RC) is coined with creating a reservoir model that contains all the attributes of a reservoir that are required to store and produce hydrocarbons. Due to non-linear and heterogeneous subterranean features, it is a difficult challenge that requires a number of sophisticated activities, including data fusion, data mining, inFormation base formulation, and uncertainty management (Anifowose et al., 2017).

Reservoir characterization is challenging because of the nonlinear and heterogeneous physical properties of the subsurface. The use of sophisticated statistical, machine learning, and pattern recognition approaches to solve such challenges has piqued the interest of researchers in the oil and gas industry. The aim of this research is to find a suitable area for drilling a new well. Sometimes, the properties associated with a reservoir system are not uniformly and linearly distributed spatially (Chaki, 2015).

The lithological parameters cannot be directly measured, they must be calculated from other geophysical logging or seismic properties. This procedure also necessitates expert participation on a regular basis to fine-tune the prediction results (Anifowose et al., 2017). Due to the significant degree of unexpected nonlinearity, conventional regression methods are ineffective for this issue. It's also crucial to understand how 3D seismic data relates to productivity, well log data, geology, and lithology. It is suggested that combining 3D seismic data with well logs can produce a better understanding of reservoir features when extrapolated out from existing wells. Improving production rates from naturally occurring complicated reservoir systems is among the oil and gas industry's most challenging tasks. As a result, identifying the patterns of the typical distributions of the relevant reservoir characteristics in the subsurface is essential (Chaki et al., 2018).

1.1 Research Issues

According to the above-mentioned literature review, the following challenges for correct reservoir characterization exist:

1.1.1 Dataset integration

Prior to modeling and categorization of lithological parameters, prepare the master dataset by combining inFormation from various sources. Well logs and seismic characteristics, for instance, are acquired using several methods with varying sampling rates and resolutions (Nwachukwu, 2018). The issue of non-unique well log and seismic data sampling, as well as variable scales of seismic, well logs, and other reservoir data, should be effectively addressed by establishing universal approaches that are independent of the targeted reservoir attribute (Anifowose et al., 2017).

1.1.2 Thin reservoir units

Changes in the sand/shale proportion are difficult to detect in areas with thin-bedded reservoirs (sand/shale units) (Nwachukwu, 2018).

1.1.3 Improper data quality

Reservoir characterization cannot be carried out using data from a research area with poor data quality or a small number of well controls and seismic coverage. With a bad data set, it's difficult to develop prediction or classification models. To account for this, pre-processing procedures will need to be fine-tuned. Uncertainties associated with the obtained dataset also led to a model's poor efficiency (Anifowose et al., 2017).

1.1.4 InFormation content

When developing a machine learning method to predict lithological features from seismic inputs, one of the most difficult issues to overcome is the predictor variables' inFormation content. If the data of the predictor variables is lower than that of the petrophysical properties, and inFormation theory-based trade-off between the quantity of inFormation needed and the actual amount of retrieval achievable must be made (Anifowose et al., 2017).

1.2 Machine Learning

Machine learning (ML), has been largely effective in commercial areas and has an enormous ability to solve problems in earth sciences as we enter the era of big data. Earth sciences applications, on the other hand, present new obstacles for machine learning because of the various geoscience aspects that are met in each problem, necessitating new machine learning research. Machine learning (ML) helps to solve obstacles posed by geoscience problems, as well as the opportunity for both ML and earth sciences to advance (Nwachukwu, 2018).

1.2.1 Implications of Machine Learning in Geosciences

Machine learning approaches offer at least 2 significant advantages over other techniques: first, they have discovered responses to questions that were hard to answer, and secondly, they can solve particular problems very quickly. Machines have demonstrated their ability to learn and utilize complex models in data or parameter-data correlations to answer problems. "Machine Learning" is a broad phrase that refers to a variety of approaches that include neural network models, support Vector Machines, variational inference, and others.

Different machines have learned to understand combinations of photos from several cameras and inFormation from a range of other instruments in a matter of seconds, allowing self-driving cars to operate in complicated, real-world surroundings (Fridman et al., 2019). Machines have learned to forecasts earthquakes (Hulbert et al., 2020), create subsurface imagery (Meier et al., 2007), and define geological features on Mars (Palafox et al., 2016) among other implementations.

1.2.2 Earth science's complexity

Earth science problems are frequently complicated (Marjanović et al., 2011). Because environmental data is frequently non-linear and contains higher-order interactions, conventional analytics may outperform when combined with missing data, as incorrect hypothesis such as linearity are imposed on the model. Machine learning surpasses traditional predictive methods in geoscience, such as predicting climate-induced changes and delineating sedimentary facies. Scientists can examine weathering responses to climate change. Delineating geologic facies aids geoscientists in comprehending an area's geology (Camps-Valls, 2020).

1.2.3 Inaccessible data

Sometimes it becomes difficult to access, collect, and interpret geo sciences data (Thessen, 2016). For this purpose, we have to utilize different techniques in terms of collection of quality dataset. For example, the mountain area having dense vegetation cover is not mapped directly (Costa, Tavares, & de Oliveira, 2019). For this purpose, we have to take help from remote sensing techniques along with machine learning algorithms. This will solve the need to go in inaccessible areas & do the work remotely.

1.2.4 Costs of time are reduced.

Machine learning can help reduce expert effort as well as save time. Traditional geological mapping methods are labor, expense, and time heavy, especially in a large, isolated location

(Latifovic, Pouliot, & Campbell, 2018). For this purpose, we have to take help from remote sensing techniques along with machine learning algorithms to do the field mapping.

1.2.5 Consistent and bias-free Data

Machine learning also has the advantage of being consistent and bias-free relative to human work. In humans, the recency effect causes classification to favor the most recently recognized classes. When it comes to the research's labeling task, if one type of dinoflagellate appears infrequently in the samples, expert environmentalists are likely to misinterpret it. Human categorization accuracies are severely affected by systematic bias (Culverhouse et al., 2003).

1.3 Area of Research

The Zamzama field is located in the lower Indus Basin, which is perceived as a promising prospect zone because several other fields have been discovered in Sindh, including the Bhit gas field, which is located south east of Zamzama, as well as Kadanwari, Miano, and Sawan, which are located north east of Zamzama.

The Zamzama field, which covers almost 120 square kilometers, is one of Pakistan's largest gas fields. We marked the reservoir Formation after acquiring nearly 196 km of 2-dimensional seismic data in 1995 as Block no. 2667-1, which is given in the Dadu Concession (Figure.1). On a daily basis, the Zamzama field accounts for 15 percent of Pakistan's entire hydrocarbon reserves. The recoverable hydrocarbon reserves are estimated to be 1.78 Tcf gas.

The main reservoir is the Pab Formation, which can be found at depths ranging from 3500 to 3800 meters and contains predominantly dry and sweet gas. There are buried anticlines which are covered with alluvium deposits. Some of the anticlines are faulted. In Sindh province, the district of Dadu is mineral-rich. The Zamzama block, which is managed by BHP Limited, features a huge processing plant that provides a boost to economic growth thanks to its large gas reserves.

According to the Upstream Petroleum Activities Map produced in October 2014 by the Directorate General Petroleum Concessions (DGPC) in collaboration with Landmark Resources (LMKR), the Zamzama is further divided into three blocks: Zamzama, Zamzama-North, and Zamzama-South. The Zamzama production and development lease was awarded to BHP Billiton, which rests with the PPL, Hycarbex owned the exploration licenses for Zamzama-North and Zamzama-South (PPL) (PPIS Map, 2014).

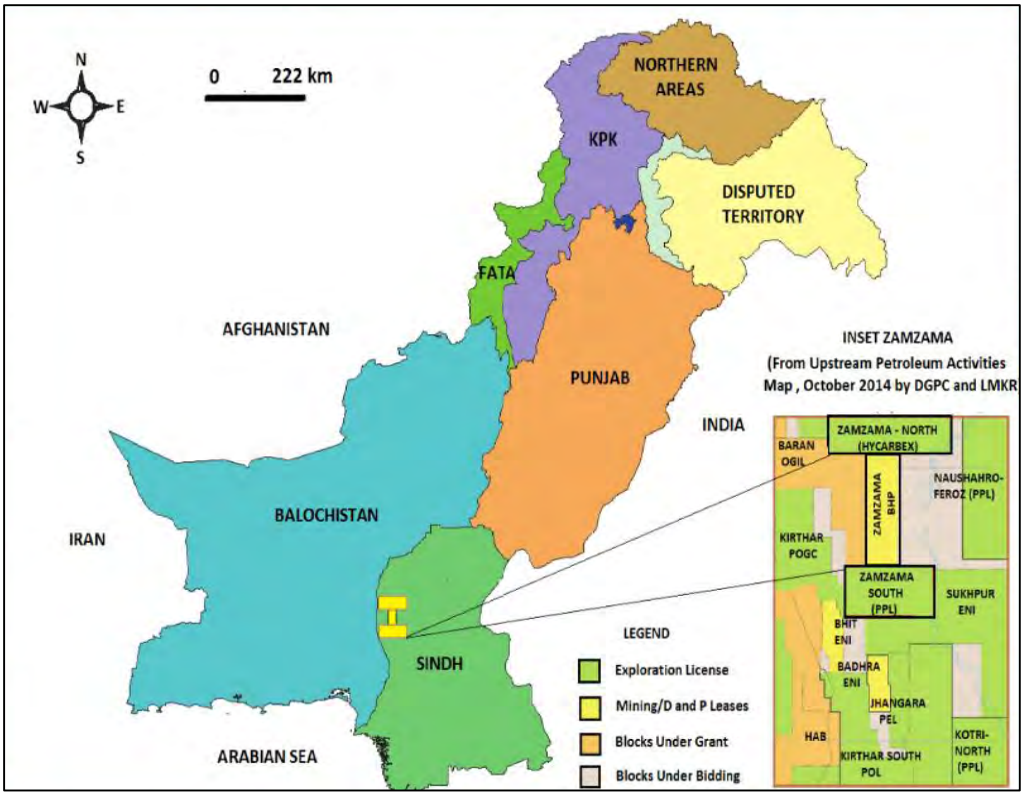


Figure 1.1 Locality of Zamzama Gas Field showing division of Zamzama blocks (Courtesy: DGPC & LMKR).

1.4 Geographical Limits

The Zamzama field is surrounded by the Kirthar range in the west, which consists of a 560 km long and 130 to 220 kilometers wide belt with a North-south direction, containing basins and valleys, while Sukkur is in the North-East and district Hyderabad is in the South-East, the river Indus is in the east, and Karachi is in the south. Lake Manchor is located in the south near the Indus River, which flows east (Figure 1.2).

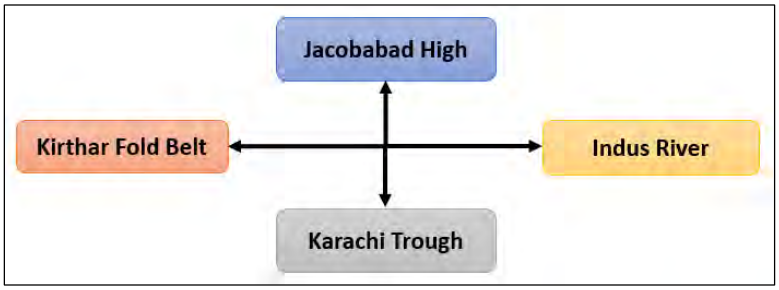


Figure 1.2: Geographical boundaries of Zamzama Gas Field

1.5 Objectives

The major objectives of the present research work are as follows.

- To check the applicability of machine learning techniques in the field of exploration geosciences.
- To test the machine learning algorithms like Random Forest, Extreme Gradient Boosting and Decision Trees for the prediction of missing petrophysical information from the available data.
- To check and compare the accuracy of each method of petrophysical data prediction.
- To perform the K-means cluster analysis for the facies classification in the Zamzama field area.

Chapter 2

2 GEOLOGY OF STUDY AREA

2.1 Introduction

Zamzama block is located in Kirthar Sub Basin which is a sub part of Southern Indus Basin which is further part of the Lower Indus Basin which is further division of Indus Basin. Geology depicts origin, structural style, mode of deformation and depositional environment. In order to understand a complete geology of study area, three parameters must be studied like Formation of basin tectonics, depositional environment sequences and modification of basin tectonics (Kingston et al., 1983). Geology of any area have an important play role in the interpretation of seismic and stratigraphic Formation (Bacon et al., 2003). The geographical and Basinal location, petroleum history, tectonic, geological setting, depositional environment, structural style, stratigraphy, petroleum system and of Zamzama block which are discussed systematically in the following.

2.2 Structural and Tectonic Setting

Pakistan is the result of Tertiary convergence zone, which formed by the interaction of 3 lithospheric Plates. These 3 lithospheric Plates Indo-Pak, Eurasian and Arabic Plates have a triple junction at the northwest of Karachi. It has large sedimentary area proven by 8 petroleum potential regions (Kazmi and Jan 1997). The Karakoram, Kohistan-Ladakh, Kharan, Chagai and Makran regions of Pakistan are comprised by Tethyan. While the Cholistan, Eastern Ranges of Balochistan and desert of Thar belongs to Gondwanian domain which is southern part of Pakistan (Kazmi and Jan, 1997). Northward movement of Indian Plate accompanied by the anticlockwise rotation, resulting sea floor spreading and collisions of Indian Plate against the Eurasian Plate were major tectonic events which give rise to local tectonics and influenced sedimentation in the sub basin.

Pakistan is being divided into these tectonic zones which are Northwest Himalayan fold and thrust belt, Kakar Khorasan flysch basin and Makran accretionary zone, Indus Platform and Fore Deep, Kohistan–Ladakh magmatic arc, East Baluchistan Fold and Thrust Belt, Chagai magmatic arc and Pakistan Offshore. Within each zone there is difference in tectonic setting and changes in structures style.

This basin where our study area lies is in extensional regime due to rifting of the Indo-Pakistan Plate. Stretching in the lithospheric Plate occurs due to rifting and as a result of thinning of the

Indian Plate. Separation doesn't turn out in beginning but instead stop stretching and cooled of edges and then subsided. This subsided region becomes a broad basin for sediment accumulation. Tectonic settings of Zamzama block are shown in Figure 2.1.

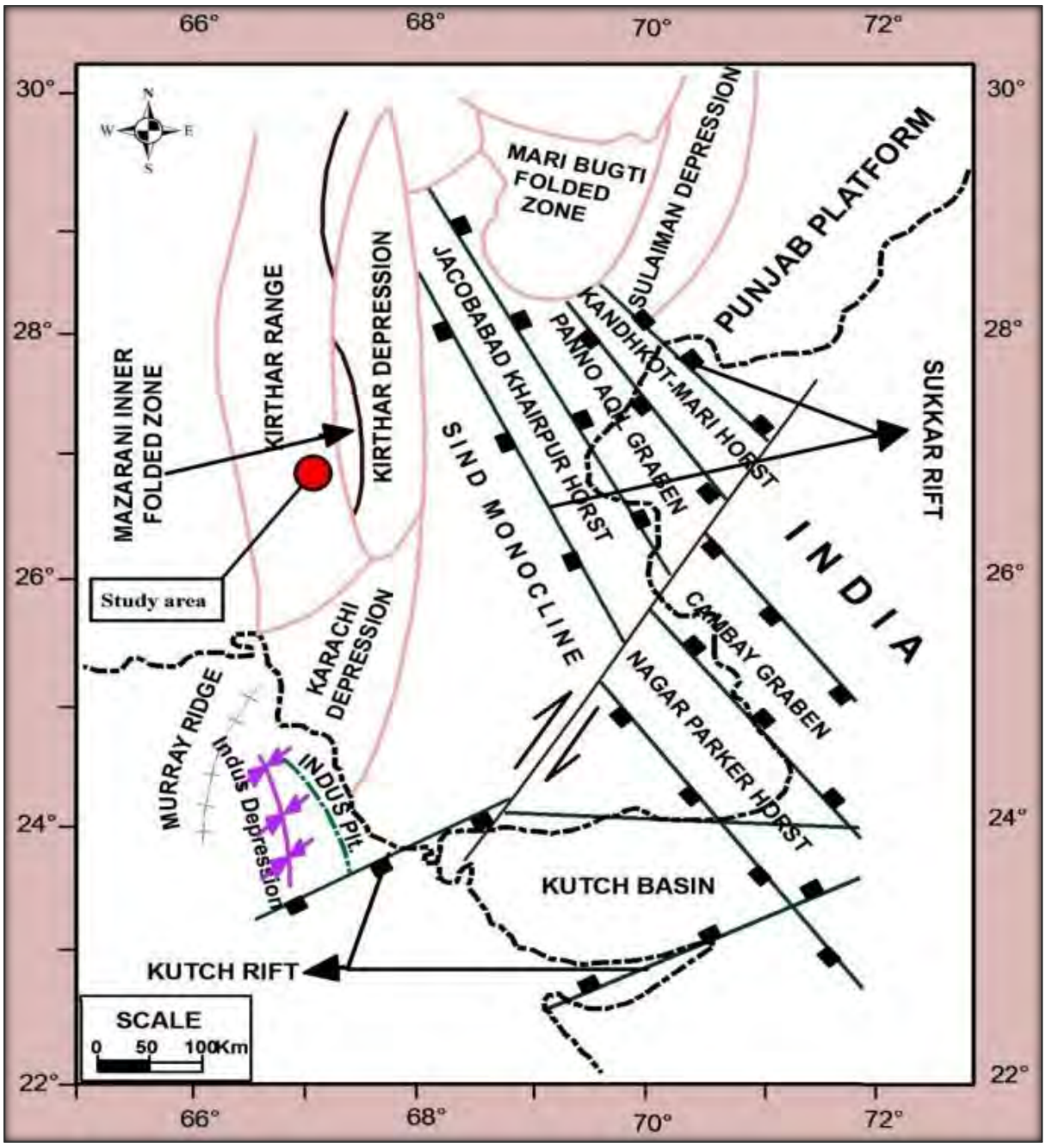


Figure 2.1 Tectonic map of Zamzama Block with red dot showing the study area(Raza et al., 1990)

2.3 Structural Pattern

The Zamzama area lies in the Kirthar Fold Belt. This area is dominated by the symmetrical and open folds that formed as a result of inversion in Jurassic aged normal fault. Thrust faults formed as a result of 2 detachment faults, one of these faults is shallow detachment in mudstone of Eocene age and other is in deep detachment of Cretaceous age. The later type is interpreted as horst and grabens along the transcurrent fault system (Zaigham & Mallick, 2000).

Multiple anticlinal structures are created on the eastern side of Kirther Fold Belt caused due to the collision during Oligocene-Miocene time (Ahmed. & Ali, 1991). The studies explain structural style and evolution of Zamzama Block that the shortening estimated along the Zamzama structure shown in Figure 2.2, describe deFormation of fault propagation folds. This Figure shows us that there is decollement under the structure, and here Zamzama contains thin structure and basement is not involved in the deFormation

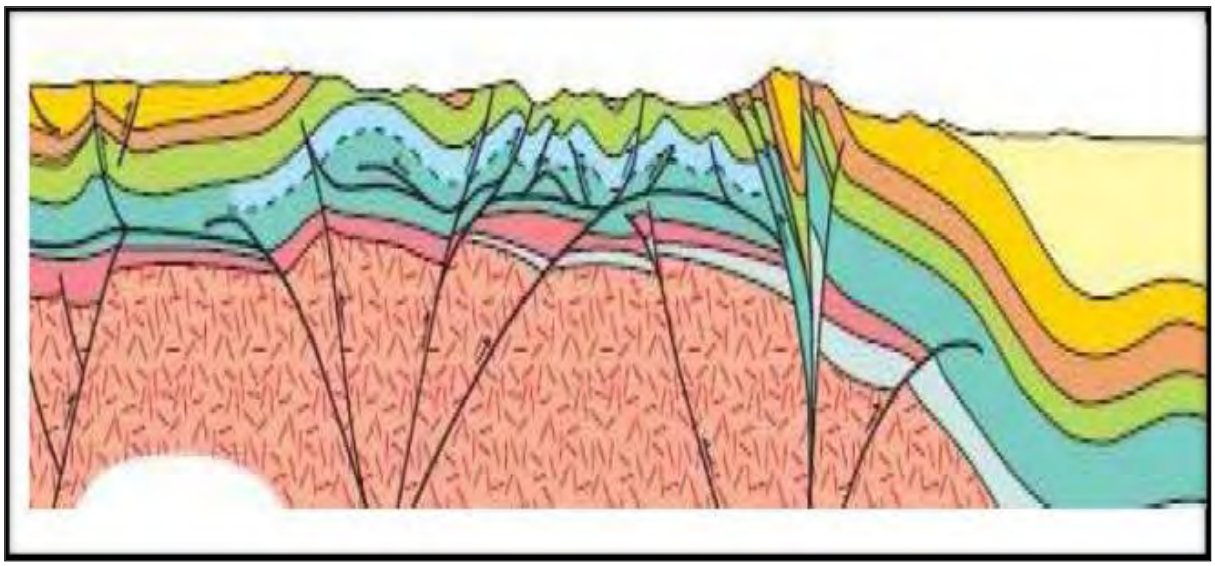


Figure 2.2 Structural pattern of Zamzama area. (From Abbasi et al. 2016)

2.4 Geological Setting

Geological setting of a sedimentary basin controls hydrocarbons and migration and entrapment. According to (Kazmi and Jan geology of Pakistan; 1997) is divided in to two regions Gondwanian domain and Tethyan domain. The southern part belongs to Gondwana and is sustained by Indo-Pak tectonic Plate. In this area rocks' age range from Triassic to Recent. The Formations comprise of sand and carbonate bodies formed in diverse shallow marine environments, ranging from shore face to lower shelf marine conditions. A large influence of sedimentary material from the northern

forces the sea to retreat to the southern side during Cenozoic age. When Paleocene period ended, Indus Basin was filled with sediments, and they have resemblance with vast flood plains and braided river system. In this area the only elevated area was the hills of Fold Belt (Kazmi and Jan geology of Pakistan; 1997).

The Lower Indus Basin is identified as an extension basin resulting from drifting and rifting of Indian Plate during Cretaceous. In Tertiary to Mesozoic sedimentary section indicates adequate source and potential reservoir and cap rocks (Kazmi and Jan geology of Pakistan; 1997).

The Lower Indus Basin is mostly comprised of Horst and Graben structures and transcurrent faults. While the fold and thrust belt is deformed due to seismic events and basement features on the western part of the basin along with the segments of the inferred rift (Zaigham and Mallick, 2000). The discoveries which are mostly in Paleocene sediments are at southern part of Indus Basin, except the Bhit gas discovery where gas is in Cretaceous sediments (Zaigham and Mallick, 2000). Due to collision and counterclockwise rotation of Indian-Eurasian Plates resulted in the Tertiary faults. There is final modification in traps and secondary migration of H.C which results in the reservoir charge. Middle Indus Basin is commonly characterized by faulted anticline structures (Kadri, 1995) (Figure 2.3).

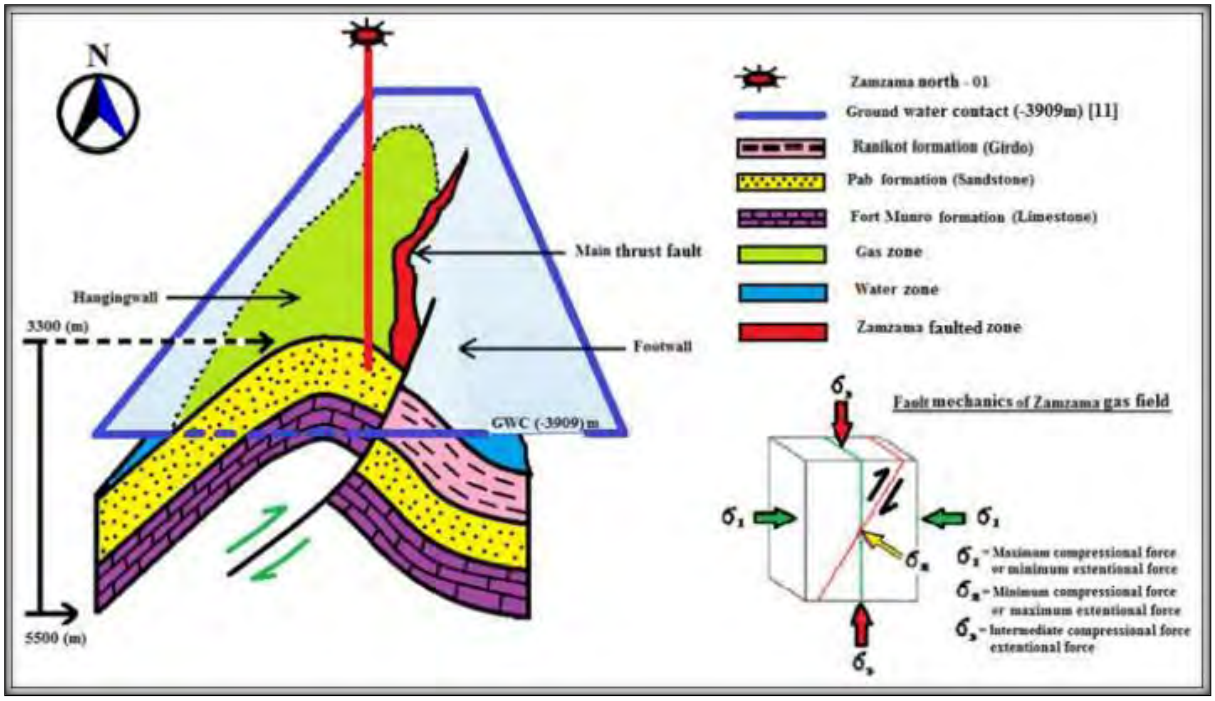


Figure 2.3 Geological structural cross section on basis of interpretation (S. Ahmed Abbasi, 2016).

2.5 Stratigraphy of Study Area

Rocks from Triassic to Recent age were deposited in Kirthar Fold Belt (Qadri, 1995). The configuration of Kirthar Fold Belt also mark the closure of Oligocene and Miocene succession (Qadri, 1995). The loasse for deep of Oligocene to Quaternary age overlying with Mesozoic to Eocene passive margin clastics and carbonate deposits that penetrate the Tertiary orogeneses of Kirthar Fold Belt are present in Lower Indus Basin along with oldest Tertiary age rocks. Rocks from Triassic to recent age were deposited in Kirthar Fold Belt (Qadri, 1995). Generalized stratigraphic chart of Kirthar Fold Belt is shown in Figure 2.4.

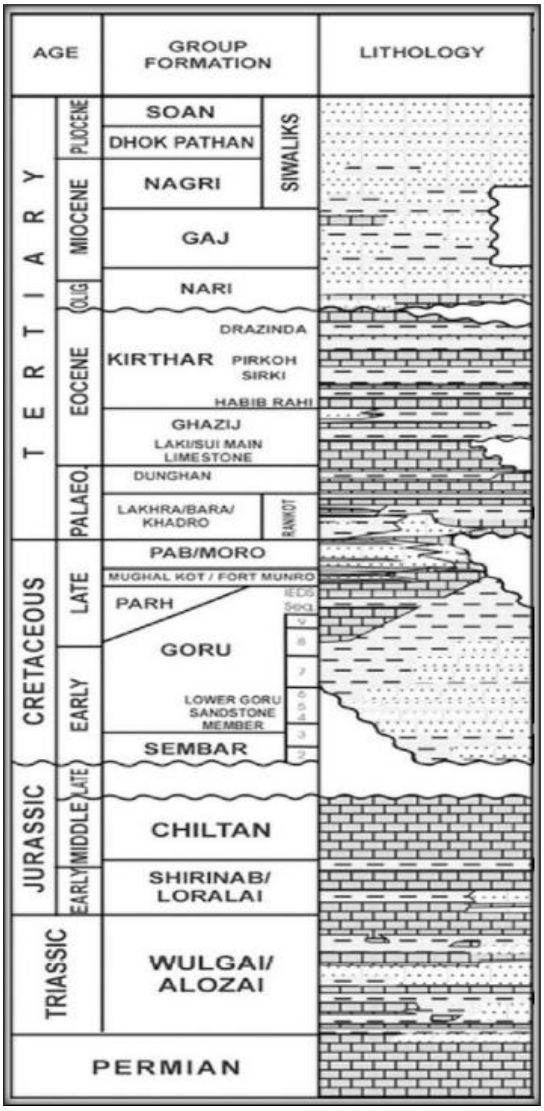


Figure 2.4 Stratigraphic succession of the LIB-Pakistan (after Zaigham and Malick 2000).

2.6 Petroleum History

In Zamzama filed first exploration well Zamzama-1/ST1, was spudded in January 1998. Extended well test (EWT) was performed in 2001 while proper production started in middle of 2003. The well was drilled up to 3,938 meters, which encountered hydrocarbons in the Khadro and Pab Formations and well logs confirmed a gas column of 300 meters. In the appraisal phase 3-D seismic acquisition done and the drilling of Zamzama-2 appraisal well. Zamzama-2 well was drilled up to 3,933 meters and also found the hydrocarbons in the Khadro and Pab Formations. Similarly, Zamzama -03, 04 and 05 were drilled as a development wells. This field is a major source of gas which is nationally ranked at fourth place according to gas production. Sweet and

dry gas with a low gas to condensate ratio of 6.5 barrels/MMcf. The estimated production life of Zamzama gas field is 15 to 25 years (Qureshi et al., 2020).

2.7 Petroleum System

Kirthar Fold Belt is high-volume gas condensate producer due to major presence of Zamzama Bhit, Mazarani and Sari-Hundi fields which guarantees petroleum system's presence in this tectono-stratigraphic province. In SIB the major gas discoveries are mostly from Paleocene sediments except the Bhit gas prospect where gas is in Cretaceous strata (Zaigham and Mallick 2000). Mature source rock, hydrocarbon migration, reservoir rock, and relative timing of the Formation of these elements and migration mechanisms are all geological components and mechanisms required to produce and store hydrocarbons.

2.7.1 Source Rock

The Cretaceous shales of Sember and Goru Formation are the source Formations of Zamzama Field. Goru Formation shows the pelagic environment whereas the Sember Formation deposited on a shelf margin. The Sember Formation deposits are mostly of marine environment. The Sember Formation has a TOC value ranging from 0.5 to 3.5 having an average of about 1.4 percent. The Sember Formation ranging from immature to over mature and have Type-III Kerogen (Ahmad et al., 2004).

2.7.2 Reservoir Rock

The Pab Formation of Late Cretaceous is the major reservoir in the Kirthar Fold Belt, and is present also in Zamzama and Bhit gas fields. It is deposited in fluvio-tidal to shallow marine environment. The Pab sandstone comprises of an extremely sand rich braid delta/coastal plain deposited system. Upper Cretaceous Pab sandstone also act as a reservoir in the Sui field.

2.7.3 Cap rock

Shales of the Khadro and Ranikot Formations of Paleocene age act as top seal for underlying Pab sandstone (Ahmad et al., 2016). These shales are proven seal in Bhit, Zamzama and Mehar fields etc. In some cases, Khadro is sandy and do not act as an effective seal but Girdo Formation act as seal (Ahmad et al., 2016).

2.7.4 Trap/Structure

Large N-S trending thrusted anticline act as trap in study area (Jackson et al., 2007). Gas was extracted from both side of fault. There is no vertical leakage, but cross wall leakage is present (Ahmad et al., 2016). The bounding fault do not reach to surface and act as seal component in Zamzama area (Jackson et al., 2007).

Chapter 3

3 DATA AND METHODOLOGY

3.1 Introduction

Reservoir characterization refers to the quantification of important effective reservoir properties such as shale volume, porosity, water and hydrocarbon saturation. Petrophysics is used for the characterization of physical reservoir parameters. This study enables to establish a record of the geological Formations in a borehole and also facilitates to quantification and identification of fluid in a reservoir (Ali et al., 2014). Petrophysical analysis gives us the inside look of the well by which we can identify the hydrocarbon bearing zones, due to its higher resolution than seismic. Moreover, the advantage of using well logs is its depth domain while seismic is in time domain, so it provides better results than only using seismic data. The generalized methodology adopted for the completion of this research work is illustrated in figure 3.1.

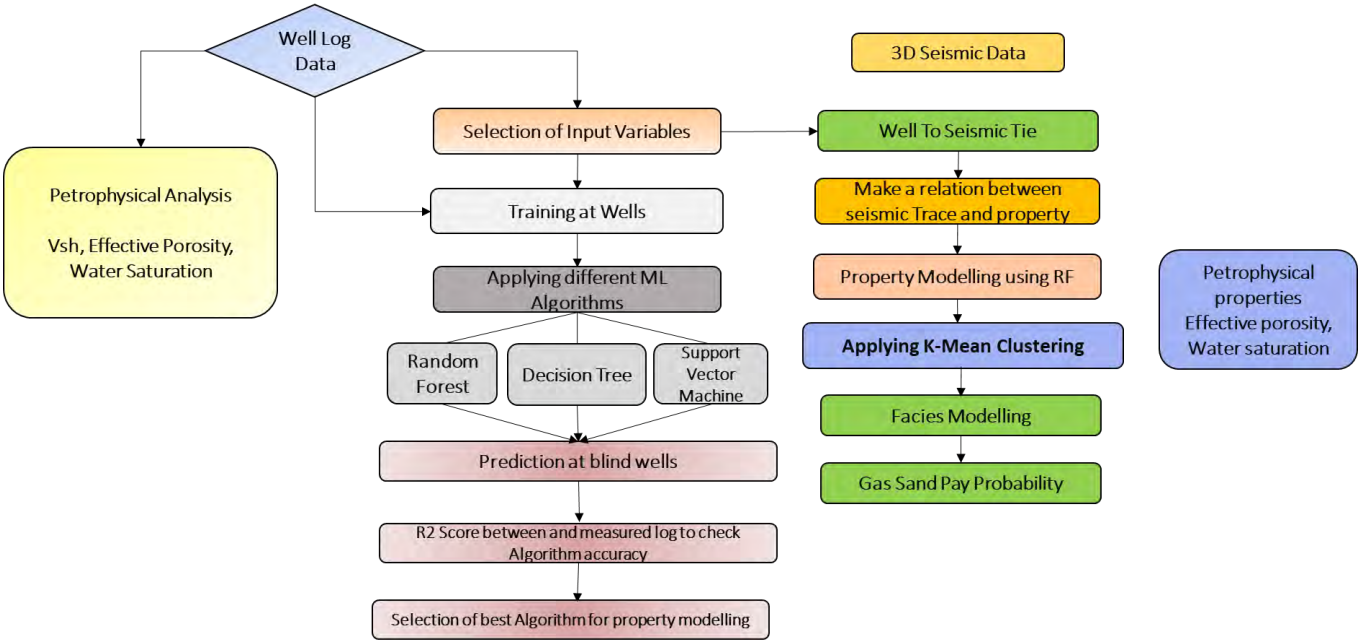


Figure 3.1. Flowchart showing the generalized methodology adopted for the research work.

3.2 Petrophysical analysis

Petrophysical analysis is carried out by performing a comprehensive study on a different wireline logs data to identify payable reservoir zone (Azeem et al., 2017). Important physical properties

like water saturation, porosity calculations, and permeability estimated through resistivity, neutron and density logs (Rider, 2002). The main objective of well log interpretation is to confirm the marked lead as a prospect by estimating a physical property of rock. So, the valuable extracted inFormation through log data helps to;

- Identification of reservoir.
- Demarcation of payable reservoir zone by applying a constraint on rock properties.
- Volumetric reserve estimation

3.2.1 Lithology Track

Gamma Ray Log, SP Log and Caliper Log are used for the lithology identification. GR log is considered as a best lithology indicator. It identifies and measures the radioactivity in rock Formation. That radioactivity mainly arises due presence of radioactive minerals such as Uranium, Thorium and Potassium. It separates shale from clean Formation (sand) (Rider, 2012).

- Gamma Ray log
- Caliper log
- Spontaneous potential log

3.2.2 Resistivity Track

It plays an important role for identification of hydrocarbon. MSFL measures the resistivity of flushed zone, LLS measures resistivity of transition zone (invaded zone) and LLD measures the resistivity of Formation's fluid in uninvaded zone. Hydrocarbon is more resistive as compared to water so, we get high resistivity values and separation between LLD and LLS against reservoir Formation. Porous sandstone shows low resistivity values due to invasion while compact and cemented sandstone shows high resistivity values (Rider, 1995). Different types of resistivity logs are:

- Shallow Laterolog (LLS)
- Deep Laterolog (LLD)
- Micro Spherically Focused Log (MSFL)

3.2.3 Porosity Track

It is defined as ratio of volume of voids to the total rock volume or measure of void (empty) spaces in a rock (Tiab and Donaldson, 2004). It is one of the important elements of the reservoir (Mavko et al, 2009). Porosities commonly ranges from 5-50 % in the reservoir rocks depending on sorting,

cementation and size of grain. Total porosity is calculated through sonic, density and neutron log or by combination of these logs. For reservoir evaluation, effective porosity is estimated by subtracting the effect of shale (Rider, 1995). Different types of porosity logs are:

- Sonic log
- Density Log
- Neutron log

3.2.4 Objective

Using well data, a petrophysical analysis was performed to determine the reservoir nature of the Zamzama region. The logs defined above will be used in order to calculate the reservoir parameters such as:

- Volume of shale (Vsh)
- Porosities (PHID, PHIT, PHIE)
- Water saturation (Sw)
- Hydrocarbon Saturation (H.C)
- Net Reservoir or Net PayVolume of Shale

In order to separate the clean Formation from dirty Formation, volume of shale (Vsh) is estimated through GR log by using a linear formula given as (Tiab and Donaldson, 2004). Volume of shale can also be estimated through Spontaneous potential log, Neutron and Density log. Caliper log evaluates the borehole geometry and identifies a lithology. Different conditions of log pattern used for lithology discrimination such as straight log pattern shows that clean reservoir Formation.

We have two methods:

a. Linear Method

In linear method we compute IGR by following formula.

$$IGR = \frac{\text{GR log} - \text{GR min}}{GR\,\text{max} - GR\ min} \quad \text{..... equation 1}$$

where, IGR is Index Gamma Ray, GR log is the recorded values of GR, GR min is minimum value of Gamma ray log, GR max is maximum value of Gamma ray log.

IGR can give us maximum volume of shale and we have to found minimum volume of shale by non-linear method.

b. Non-Linear Method

In non-linear method we have various formulas like Stabier, Larinov and Clavier to compute minimum volume of shale. We utilize the one which give us minimum volume of shale. And mostly Stabier give us minimum volume of shale.

- **Stabier:** (Most preferable)

$$Vsh = \frac{\text{IGR}}{3-2\,\text{IGR}} \quad \text{…..……} \quad \text{equation 2}$$

Where, IGR= Index Gamma Ray

- **Larinov:** (Used for Older rocks)

$$Vsh=0.33(2^{2\,IGR} - 1)$$

- **Clavier:**

$$Vsh=1.7-(3.38-(IGR+0.7)^2)^{0.5}$$

3.2.5 Porosity

In the next step we have to calculate Porosity parameters, like

- Density Porosity
- Sonic porosity
- Effective Porosity
- Neutron porosity (Given)

3.2.5.1 Density Porosity

Density log data is given but we need density porosity for the cross plot with Neutron porosity to have better interpretation. Porosity values calculated from density log is call density porosity.

$$RHOB\ \Phi = \frac{(RHOB\ mat - RHOB\ log)}{(RHOB\ mat - RHOB\ fluid)} \text{ ……equation 3}$$

where, $RHOB\ \Phi$ is density, $RHOB\ log$ is density log, $RHOB\ mat$ is value of matrix density, $RHOB\ fluid$ is density of fluid

The estimated value of density of matrix used is 2.65 gm/cm^3 which is for Sandstone and density of fluid is $1 gm/cm^3$.

3.2.5.1.1 Sonic Porosity

For Sonic porosity we will use formula of consolidated rocks because we know that these rocks are old and well consolidated.

$$\Phi s = \frac{\Delta T \log - \Delta T \text{ mat}}{\Delta T \text{ fluid} - \Delta T \text{ mat}} \quad \text{..... equation 4}$$

where, Φs is sonic porosity, ΔT log is sonic log, ΔT mat is value of travel time in matrix, ΔT $fluid$ is value of travel time in fluid. Φs is value of sonic porosity

The delay time of Formation increases due to hydrocarbon also known as hydrocarbon effect. This influence should be removed because it affects the values of calculated porosities.

3.2.5.1.2 Effective Porosity

The void spaces in the Formation that contribute to fluid movement or permeability in a reservoir. Discrete pores and water trapped on clay minerals are excluded.

Effective porosity is normally less than actual total porosity. Effective porosity log was created by using total porosity logs and volume of shale log.

The mathematical relation for effective porosity is as follows:

$$\Phi_e = (1\text{-}V_{sh})*\Phi_{avg} \quad \text{..... equation 5}$$

where, Φe is effective porosity, Φavg is average porosity, Vsh is volume of shale.

3.2.5.1.3 Neutron Porosity

Neutron log is sensitive to the hydrogen atoms present in a Formation and determination of the porosity of a Formation. Count rate will be low in high porosity rocks and vice versa.
Neutron porosity is given in the data and calculated by well log w.r.t depth.

3.2.5.1.4 Total Porosity

The total porosity is the average porosity obtained from various logs divided by the total number of logs used for calculating porosities.

$$\varphi T = \frac{\varphi d + \varphi n + \varphi s}{3} \quad \text{..... equation 6}$$

where, φT=Average porosity, φd density porosity, φn neutron porosity, φs sonic porosity

3.2.5.1.5 Saturation of Water

Water Saturation can be calculated by having resistivity of water Rw. Rw can be calculated by having various parameters like bottom hole temperature, surface temperature, water salinity in ppm and SP (Static). (Amigun et al., 2012).

Two methods are being used for resistivity of water.

1) Pickett crossplot Method
2) SP Method

3.2.5.1.6 Pickett Crossplot Method

Pickett Crossplot (Pickett, 1972) is simple and effective methods used to measure the resistivity of water, in this method we not only get the estimates of water saturation, but we also determine the following factors:

1. Formation water resistivity
2. Cementation factor
3. Matrix parameters for porosity logs

The true resistivity Rt, that is a function of porosity, water saturation, and cementation factor, is used in the Pickett technique. On2/3 cycle log-log paper, a Pickett cross plot is created by plotting porosity vs deep resistivity (LLD) data. Pickett Crossplot comprises of number of water saturation lines in which lowest or left-most line shows water bearing line with 100% saturation. Data points lies above this straight line (Ro) represent water saturation less than 100%. Slope of straight line represent cementation factor(m) water-bearing line. The generalized Pickett plot is illustrated in figure 3.2.

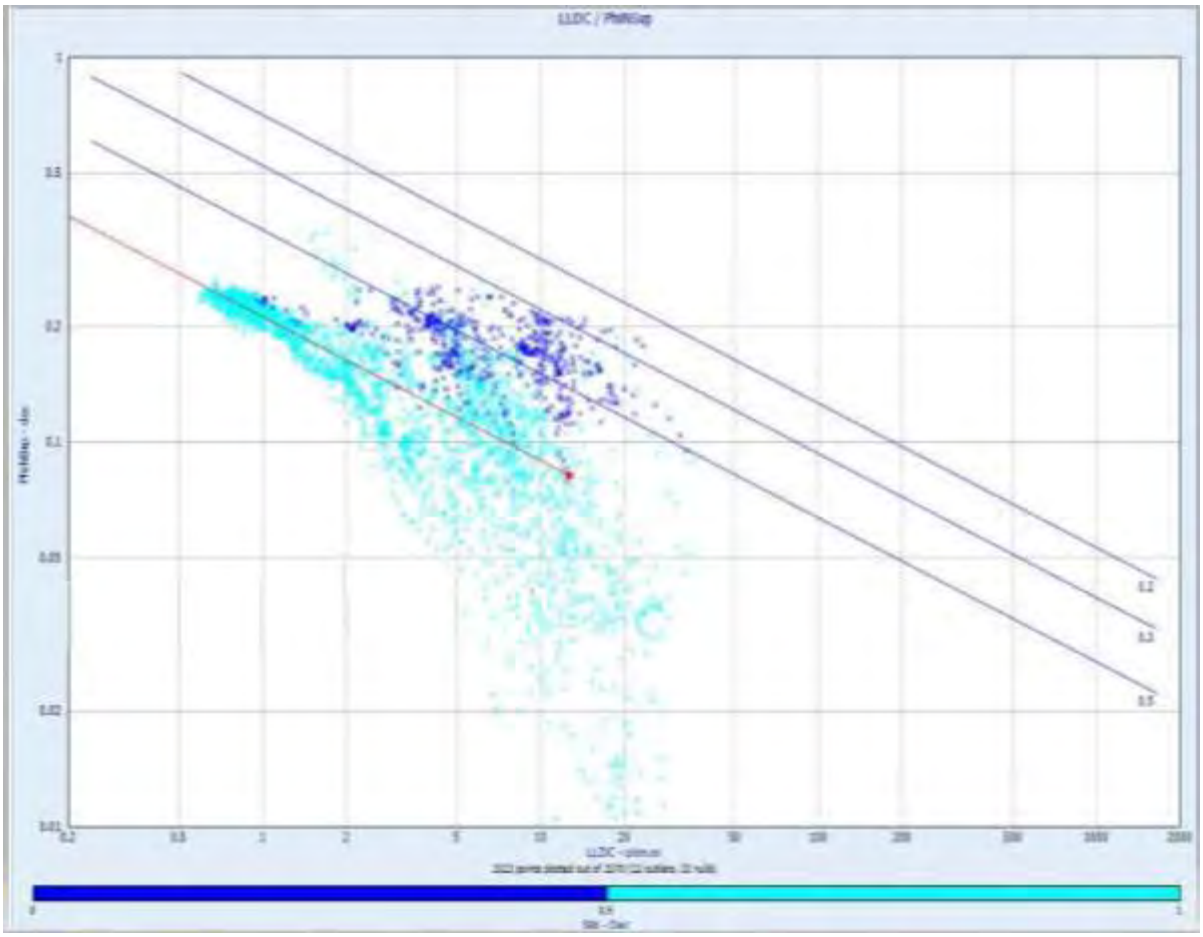


Figure 3.2. Generalized Pickett plot is shown the red line is indicating about the 100 percent saturation of water.

3.2.5.1.7 Hydrocarbon Saturation

The fraction of pore spaces containing HCs is known as hydrocarbon saturation. The simple equation used for this is given below.

$$S_w + S_H = 1$$

The saturation of HCs is percentage of pore volume occupied by hydrocarbon.

$$S_H = 1 - S_w \text{ ……. equation 7}$$

where, S_H= Hydrocarbon saturation

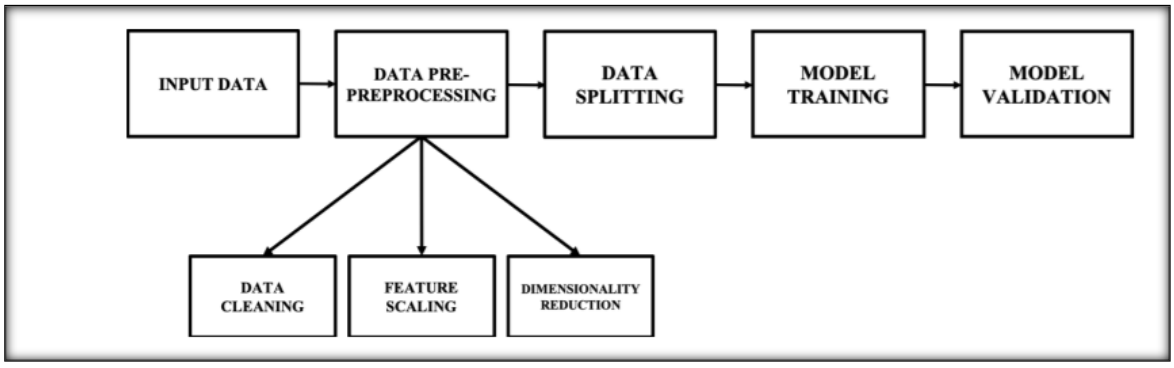
S_w= Water saturation.

3.3 Machine Learning

In artificial intelligence, machine learning is the branch of the field that can learn from data and recognize patterns in order to make judgments with little or no human interaction. Because of the abundance of easily available data, the petroleum industry is particularly well positioned to

benefit from machine learning. The use of machine learning techniques allows for the discovery of correlations between physical parameters from both input and output data, as opposed to conventional or physical modelling approaches, which require modelers to explicitly account for such associations while setting up a model. The reservoir characterization of the Zamzama field was carried out using machine learning algorithms.

Many scholars have used data-driven approaches to tackle geological issues in recent years, thanks to the use of machine learning in numerous sectors of science and engineering. Known as machine learning in the area of artificial intelligence (AI), it is defined as a training process that allows computers to learn and act by using an extensive collection of algorithmic techniques. Machine learning can be defined in a variety of ways and from a variety of perspectives. A computer learning model with which it can assess samples is defined by Nikhil (2017) as giving a computer a limited set of instructions to adapt the framework when it makes a mistake, rather than instructing the computer to solve the problem by trying to teach it a vast list of rules to address the problem. Because the fundamental aim is to make inferences from a sample, Alpaydin (2014) defines machine learning as the use of statistical concepts in the construction of mathematical models. He goes on to say that it involves utilizing example data or prior expertise to train computers to maximize a performance criterion. Once a model has been created up to a certain point, learning is the process of running a computer algorithm to optimize the model's attributes using training data or past experience once the model has been established. There are many types of models that may be used to generate forecasts in the future, descriptive models that can be used to learn from data, and models that are both.



When considering adopting machine learning to solve a specific problem, the first thing to

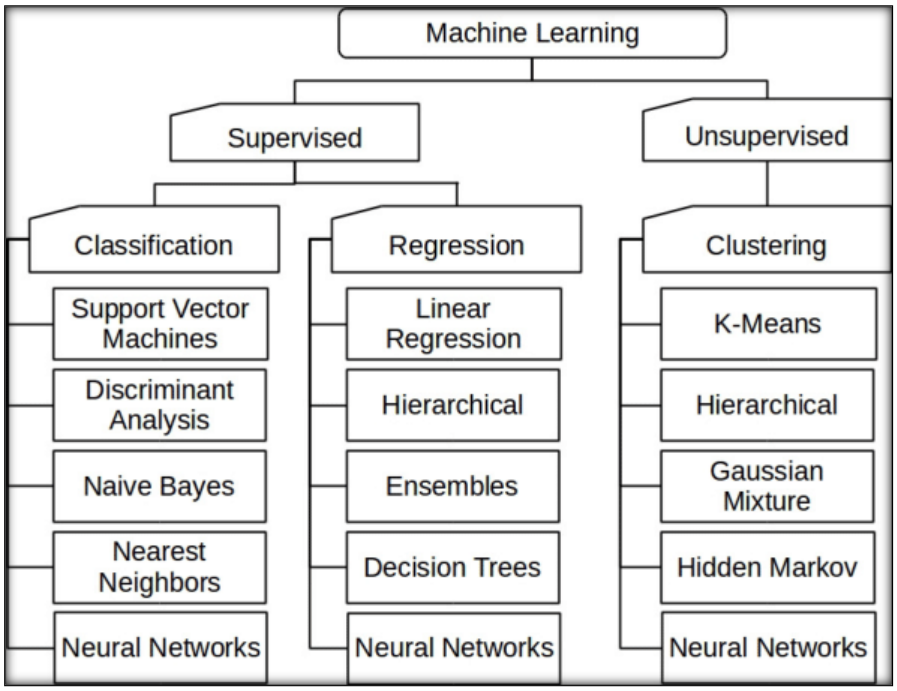
Figureure 3.3 Workflow of machine learning (From Caté et.2017)

consider is if ML is the best technique for tackling it. Furthermore, because machine learning may never execute flawlessly in real-world challenges, there are a number of factors to consider before starting an ML project. These factors include the availability of a huge quantity of data, the absence of a high level of precision, and the fact that the problem is well-understood, allowing for the development of appropriate algorithms (Awad and Khanna, 2015). As a result, after the essential prerequisites are completed, the workflow in Figure 3.1 below depicts the procedure we will use to construct the present machine-learning project. Each machine learning-based study will have its own methodology, which will alter depending on the desired findings. However, most machine learning operations, whether complicated or basic, will adhere to a set of rules.

Two basic types of machine learning are the supervised/predictive learning approach and the unsupervised/descriptive learning technique. Reinforcement learning is a third machine learning technique that is less commonly used.

3.3.1 Unsupervised Machine Learning

Unsupervised machine learning algorithms attempt to learn patterns in a dataset without the use of a user-defined label or target, i.e., without or with little supervision. The feature dataset is used as the input in unsupervised learning, while the output is determined by the algorithm (Hussain et al. 2021). Clustering is widely used technique in unsupervised machine learning that is natural grouping of data. Different algorithms are used further used in clustering that includes K-Means, Gaussian Mixture Model, Hierarchy clustering algorithm, Neural network and Hidden Markov as shown in Figure 3.2.



Figureure 3.4 Types of Machine learning technique. (From Ayodele,2010)

3.3.2 Supervised Machine learning

Machine learning algorithms that are supervised attempt to "learn" the link between the feature and the goal. This is accomplished by developing a function that translates the inputs (feature variables) to the outputs (target variable). A feature and a target dataset are used in supervised learning are necessary in order to fit and train the model (Jian et al.2020). The following model is then utilized to make predictions on an individual basis dataset of unknown/unseen features with no associated goal. The target dataset in classification is a set of discrete variables. Supervised learning techniques are frequently capable of handling both regression and classification problems; the method is normally built for one instance and changed for the other (Caté et al.2017). Linear regression, Support Vector Machine, Random Forest, and other supervised learning algorithms are examples.

In this research work, regression algorithms of supervised machine learning were used. The algorithm used were

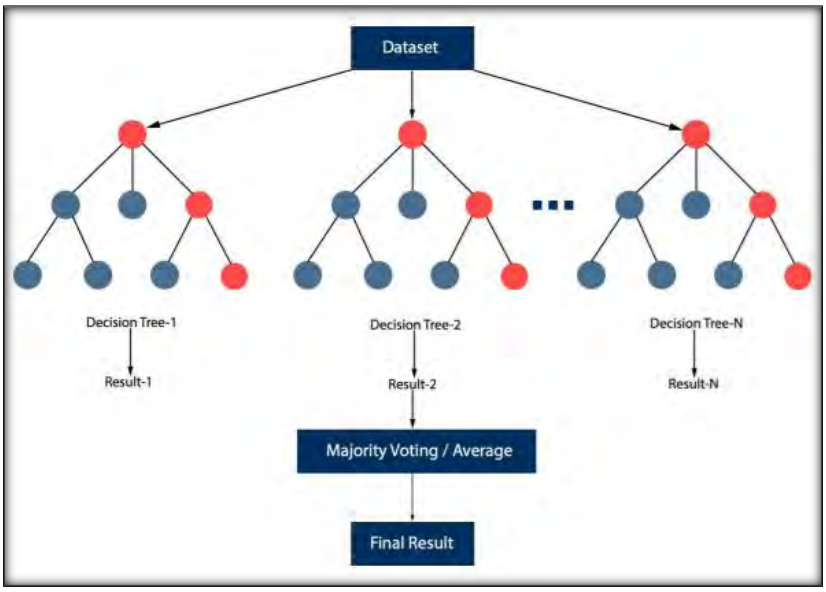
- Random forest
- Support Vector Machine
- Decision Tree

- Extreme Gradient Boosting (Xgboost)

3.3.3 Random Forest

When modeling data, random forest makes use of a large number of single decision Trees, which is known as the random forest approach (Liaw and Wiener, 2002). Each decision Tree provides a sub-model and a prediction for the situation. Random forest model assesses the result of each decision Tree's forecast and selects the prediction with the greatest probability as the model's final output. Random forest model evaluation and selection was also described by (Pal, 2005). Throughout the training process, every decision Tree is trained using a randomly selected set of data points from which it may be built. These sets of data points are picked using replacement, which implies that some data sets may be utilized more than once to train different decision Trees in different situations. In addition, the attributes that are employed in each decision Tree are picked at random by the algorithm.

This implies that each decision Tree is formed using a subset of features that has been randomly picked (Feng et al., 2021). Using a large number of randomly resulted independent decision Trees that were developed from a randomly chosen collection of data points, the random forest model lessens model bias and improves model performance by taking into account the output results from a significant number of randomly produced independent decision Trees as shown in Figure 3.3 In this example, N represents the number of training samples, and label M in Figure 3.3 represents the number of feature types. When m features are entered into the decision Tree (m M), unsampled samples are used to forecast and assess error (Feng et al., 2021). The put-back technique is used to sample N times (i.e., bootstrap sampling) to construct a training set, unsampled samples are used to forecast and assess error after the error has been calculated. Each node on the decision Tree has m attributes that are picked at random, and judgements are made for each node on the decision Tree based on these qualities. The appropriate split mode is then computed based on the m features that have been determined. If one leaf node of the decision Tree continues to split or if all samples point to the same category, the size of each Tree will ultimately grow without being pruned (Wang et al., 2020).



Figureure 3.5 Illustration of random forest algorithm structure.
From Rodriguez-Galiano et al.2015

3.3.4 Support Vector Machine

Support Vector Machines (SVM) seeks to determine the equation of a multi - dimensional surface that, from a geometrical standpoint, best separates distinct classes in the feature space by determining the equation of a multidimensional surface. In contrast to other approaches, SVM strategy resolves the convex optimization challenges analytically, which implies that it will always provide the same hyper-plane parameter regardless of how the model is started. Similarly, other commonly used algorithms for classification problems, such as the perceptron, accomplish their findings based on initialization and termination criteria, thereby constituting an iterative process (Awad and Khanna, 2015).

3.3.5 Decision Tree

In real life, a Tree structure has numerous analogues in many industries. The decision Tree algorithm may be used to visually show the decision-making phases throughout a decision-making process. In many diverse disciplines, including classification and regression issues, the decision Tree technique is widely employed in data-driven modeling. The direction of a decision Tree model is the inverse of that of a real Tree structure. The construction of a decision Tree is shown in Figure 3.4, with the root at the top (Safavian and Landgrebe, 1991). Each circle, which symbolizes an internal node, is centered on the point at which the Tree structure divides into further branches. The leaf, or decision, is the end of the Tree structure that does not divide further. In data-

driven models, they represent the various labels or classes (Friedl and Brodley, 1997). The technique uses a cost function to determine the cost of each split at each conditional node. As the root node, the node with the lowest cost is chosen. The Tree divides at the root once the root node is chosen, and the algorithm searches for nodes at lower levels.

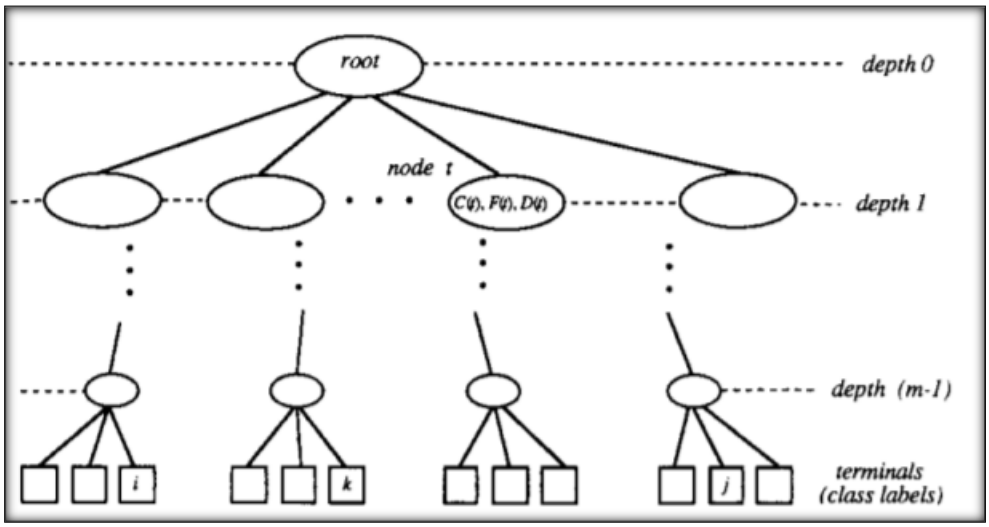


Figure 3.6 Illustration of decision Tree (Safavian and Landgrebe, 1991)

3.3.6 Extreme Gradient Boosting (Xgboost)

Gradient Boosting machines (gbm) are one of the top performing algorithms for supervised learning, and Xgboost is one of the applications of this technique. It may be used to solve problems involving regression and classification. Chen and Guestrin in 2016 proposed XGBoost, a massively scalable edge Tree boosting system that has been widely used and refined in a variety of research domains. An ensemble of classification and regression Trees is used in this approach (CART). Furthermore, by reducing a regularized objective function, this series of Trees are utilized to fit the training data. The outputs of each Tree are added together to minimize the model's total residual and achieve regression. Data scientists choose Xgboost because of its fast out-of-core compute execution speed.

3.3.7 K-Means Clustering Technique

This technique is employed for unsupervised machine learning problems that are challenging. K-means clustering aims to create a cluster of similar items in a data set on the basis of similarity (Ahmad, & Dey, 2007).

The clustering or sorting out technique is widely used in daily life too. For instance, if we go to the supermarket, we observe that beauty products are kept in one section while snacks are in another section. Apples are stacked in one cluster while carrots are in another cluster. Each product forms one cluster as it belongs to the same category.

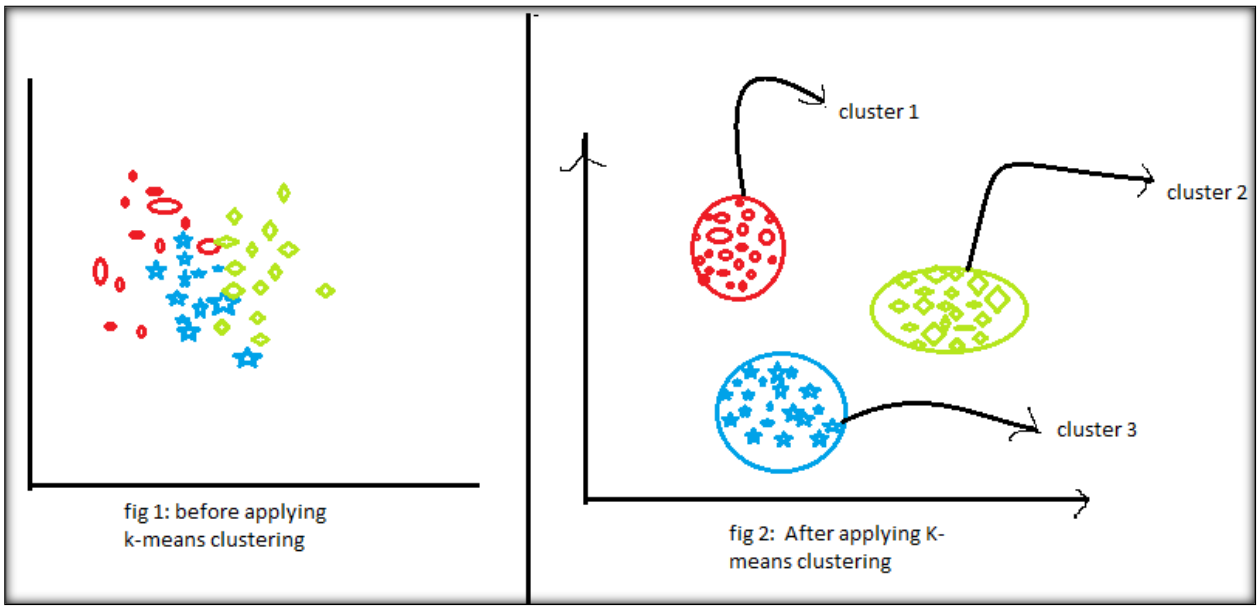


Figure 3.7 K-Means Clustering. (From Ahmad, & Dey, 2007)

Take a look at the 3.5 diagrams above. So, what did you notice? Let's have a look at the first illustration. The data is shown in the first Figure before the k-means clustering technique is applied. All 3 categories are twisted up in this case. When confronted with such facts in the physical world, you will be unable to distinguish between the many groups. This is the data after the K-means clustering method has been used. As you can see, the three separate things are divided into three distinct groups known as clusters.

3.3.7.1 Advantages of K-means

- It's easy to put into action.
- K-means clustering technique can manage large data sets and work efficiently while dealing with large data.
- This method can handle new cases.
- Clusters with various shapes and sizes can be generalized.

3.3.7.2 Drawbacks of K-means

- It is highly sensitive to outliers.
- Manually selecting the k values is a difficult task.

- Scalability reduces as the data size grows.

Chapter 4

4 RESULTS AND DISCUSSION

4.1 Petrophysical Interpretation

The main reservoir of Zamzama Gas field is Pab sandstone. By inspection of lithology, porosity, resistivity, effective porosity and water saturation (table 4.1). Separation observed between LLD and LLS. Neutron and density log show crossover. Caliper logs show a stable behavior being compact sandstone. Effective porosity is acceptable with good percentage of hydrocarbons. On basis of observed logs behavior and its calculation these zones are marked as potential zone. The petrophysical interpretation of Zamzama 03 (Figure 4.1), Zamzama 02 (Figure 4.2), Zamzama 05 (Figure 4.3), Zamzama 06 (Figure 4.4) and Zamzama 08-st (Figure 4.5) is shown.

Table 1 Calculated values of the Pab Formation at Zamzama wells

	ZZ-03	ZZ-02	ZZ-05	ZZ-06	ZZ-08-st2
VCL (%)	10-12	14-16	20-22	18-20	25-30
PHIE (%)	10-12	11-13	9-11	11-13	8-10
SW (%)	25-30	15-20	10-15	20-25	30-35
Sg (%)	70-75	80-85	85-90	75-80	65-70

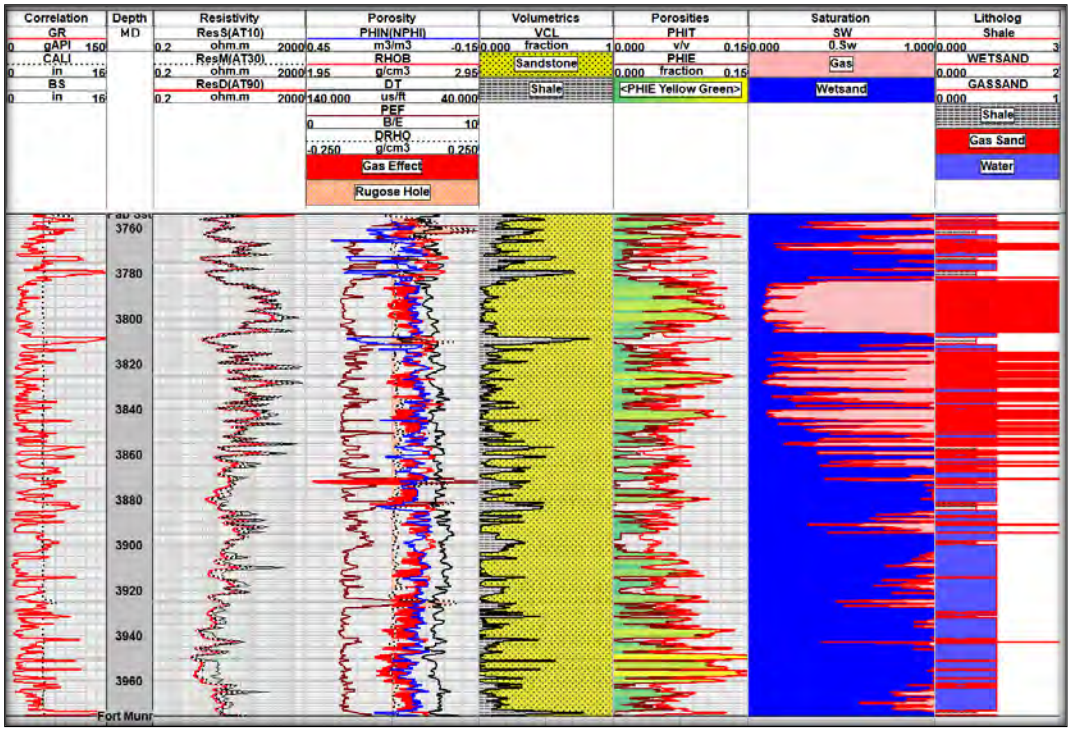


Figure 4.1 Petrophysical analysis of Zamzama 03 well

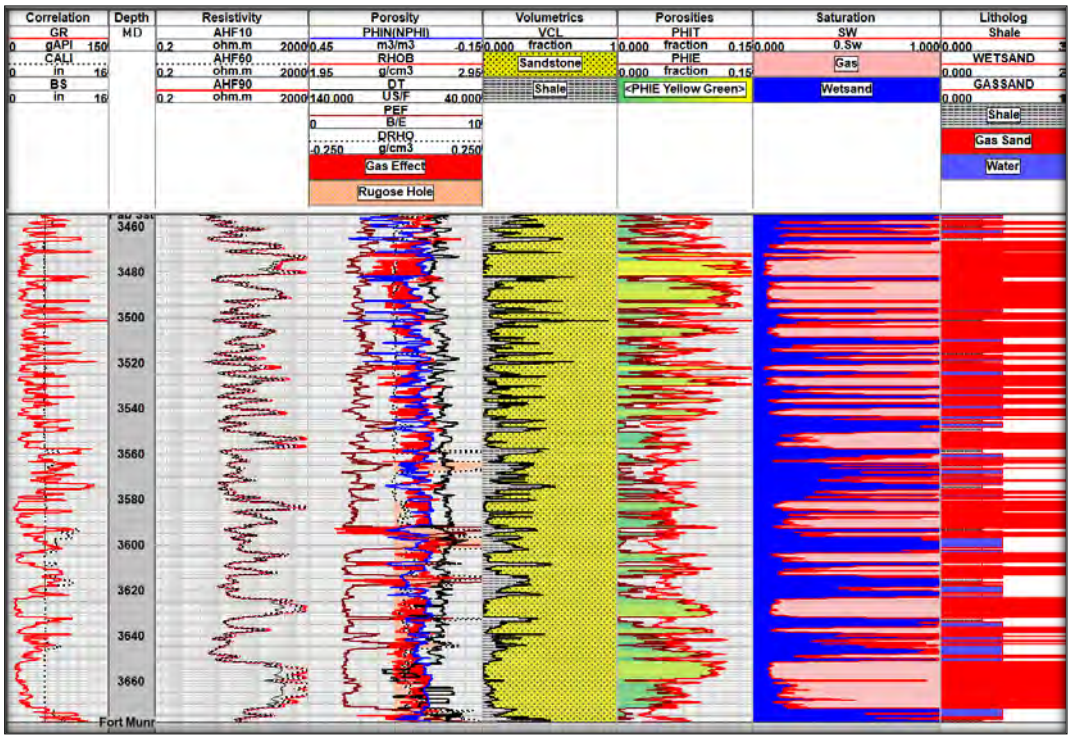


Figure 4.2 Petrophysical analysis of Zamzama 02 well

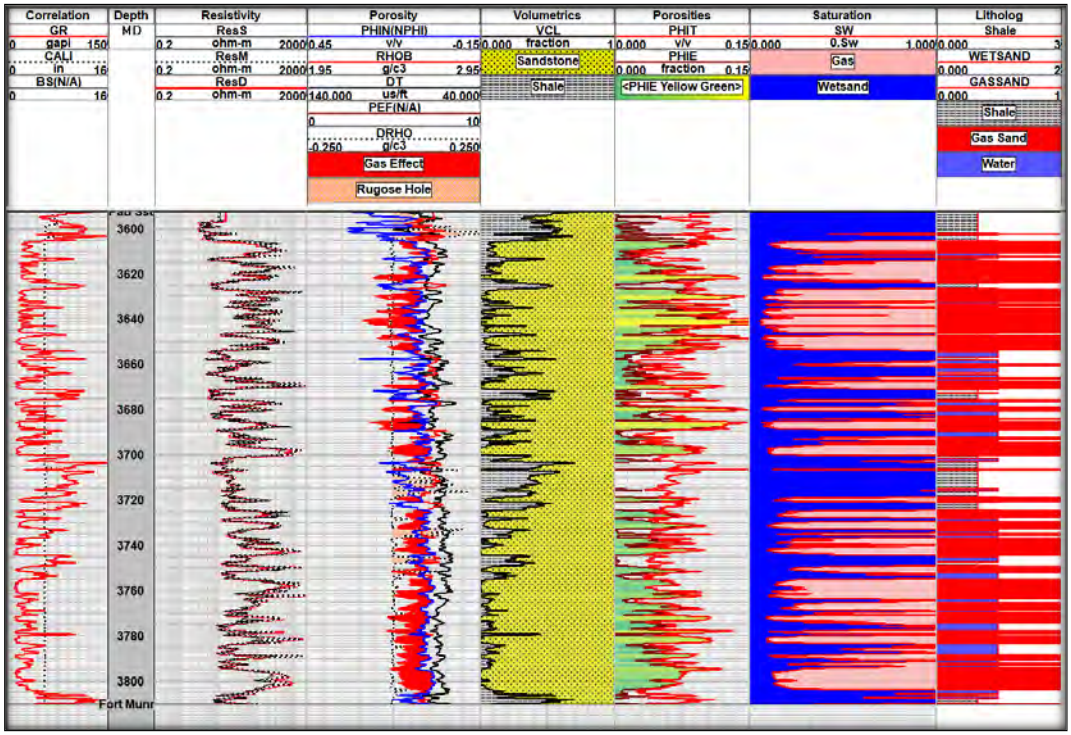


Figure 4.3 Petrophysical analysis of Zamzama 05 well

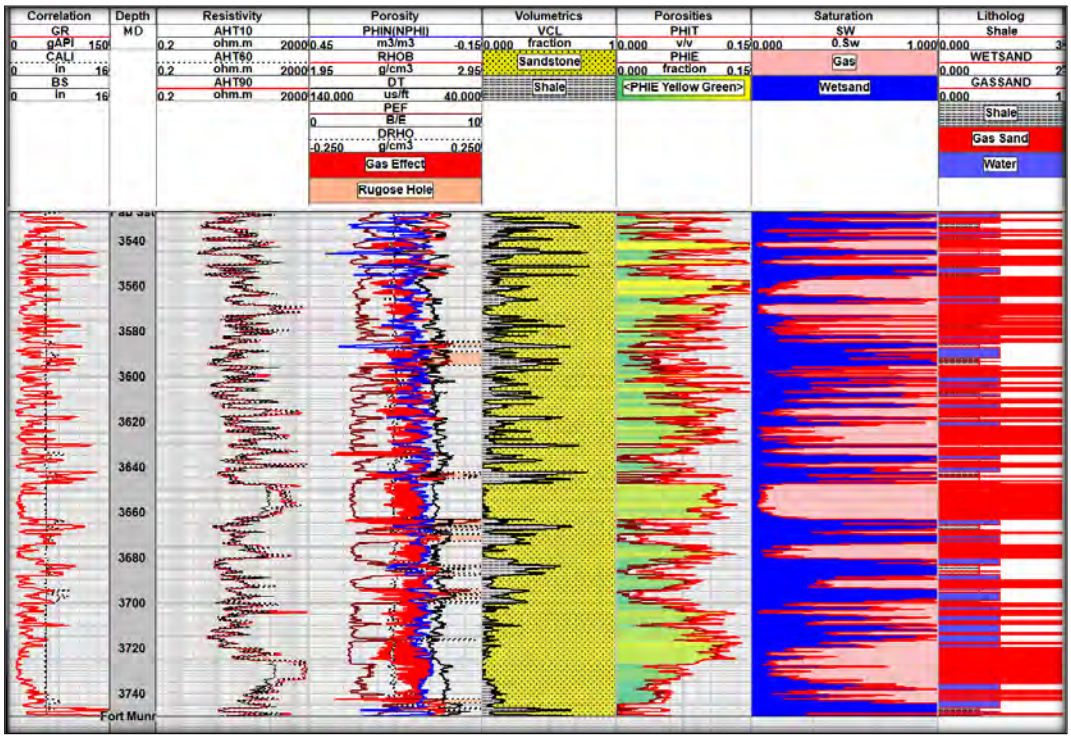


Figure 4.4 Petrophysical analysis of Zamzama 06 well

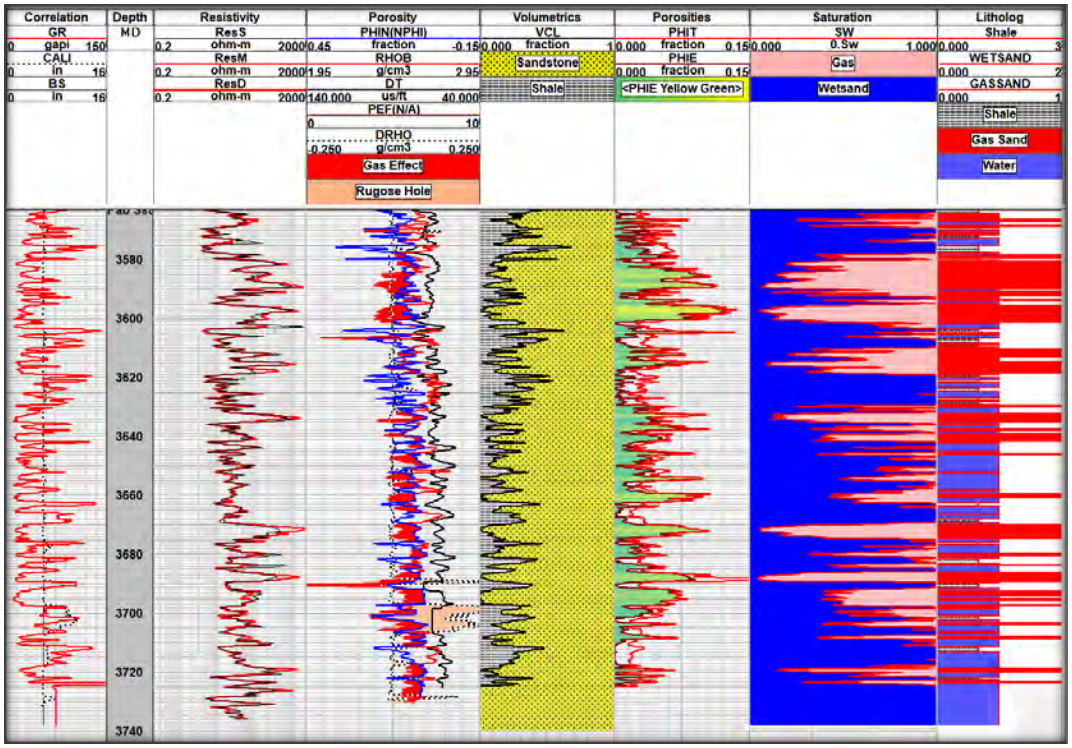


Figure 4.5 Petrophysical analysis of Zamzama 08-st-2 well

4.2 Input Dataset of Machine Learning for Zamzama Gas field

The initial data sets that are used to train machine learning models is referred to as training datasets. Machine learning algorithms make predictions or perform a desired task by using training data sets. For a study area, GR, RHOB, and DT curves of ZZ-02 and ZZ-03 shown in Figure 4.6-, ZZ-05, ZZ-06 in Figure 4.7 and ZZ-08 in Figure 4.8 are the input curves of machine learning for prediction.

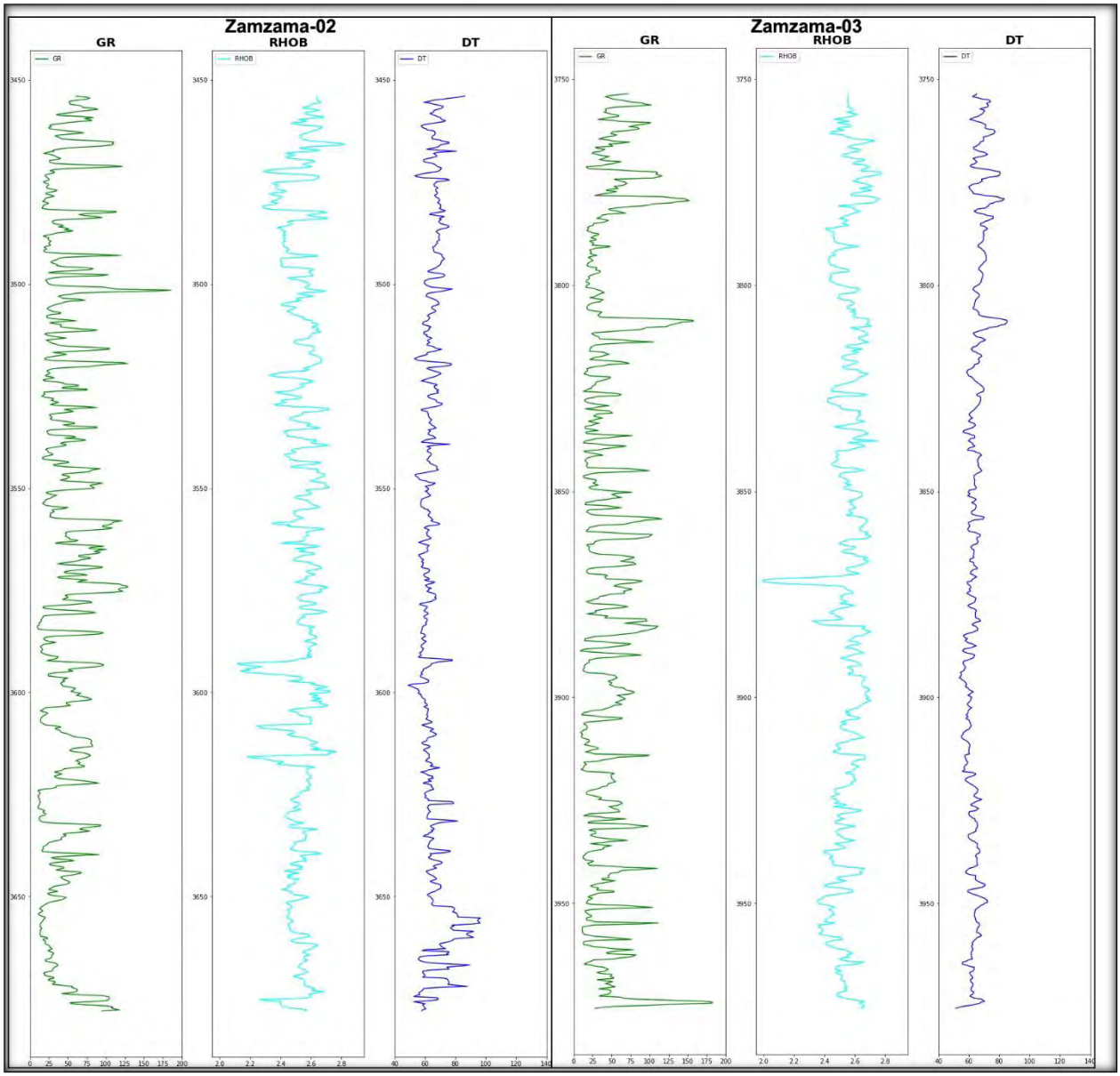


Figure 4.6 Input Dataset of machine learning for Zamzama-02 and Zamzama-03 used in training for the prediction of VCL, PHIE, SW.

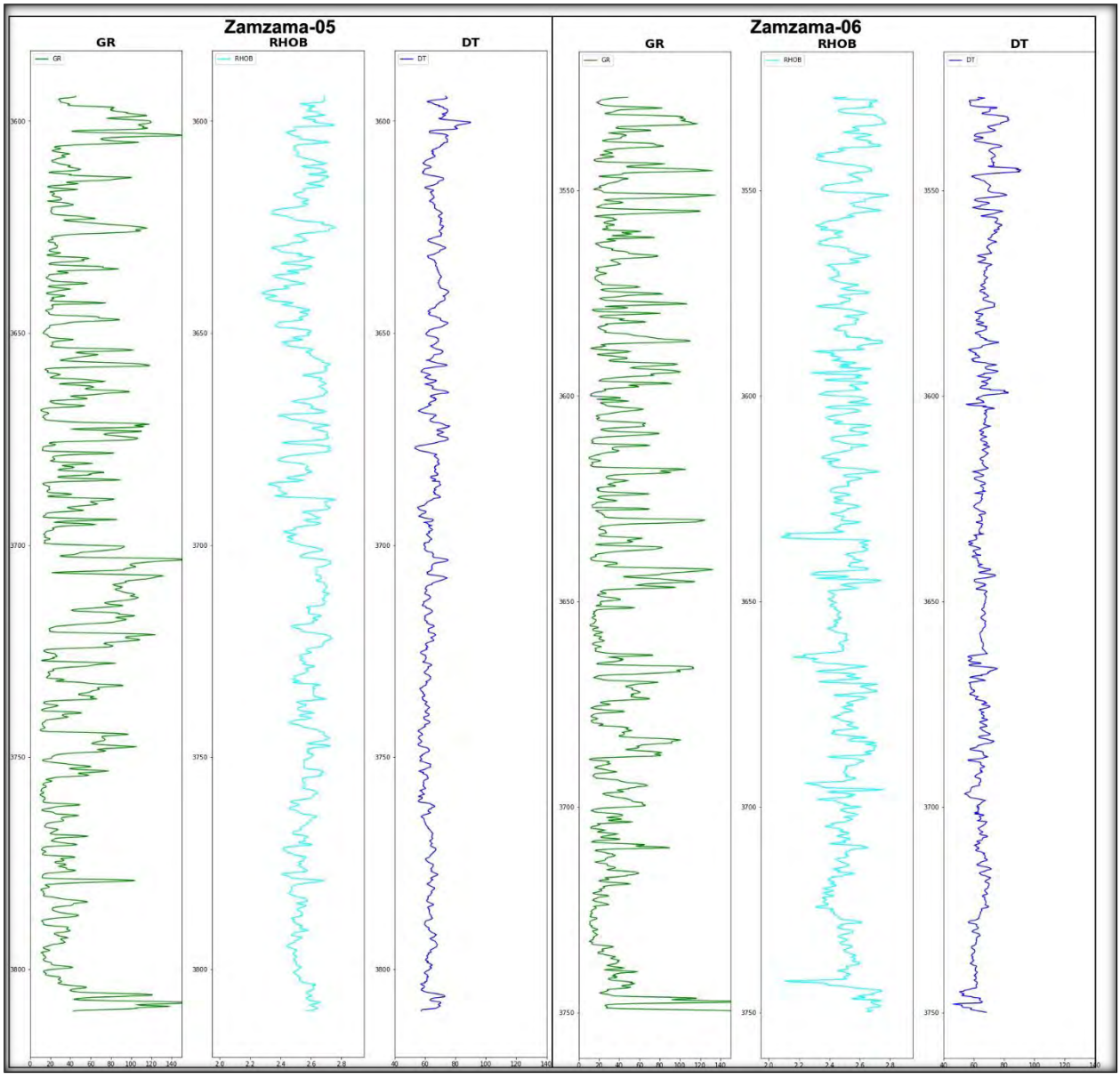


Figure 4.7 Input Dataset of machine learning for Zamzama 05 and Zamzama-08 used in testing for the prediction of VCL, PHIE, SW

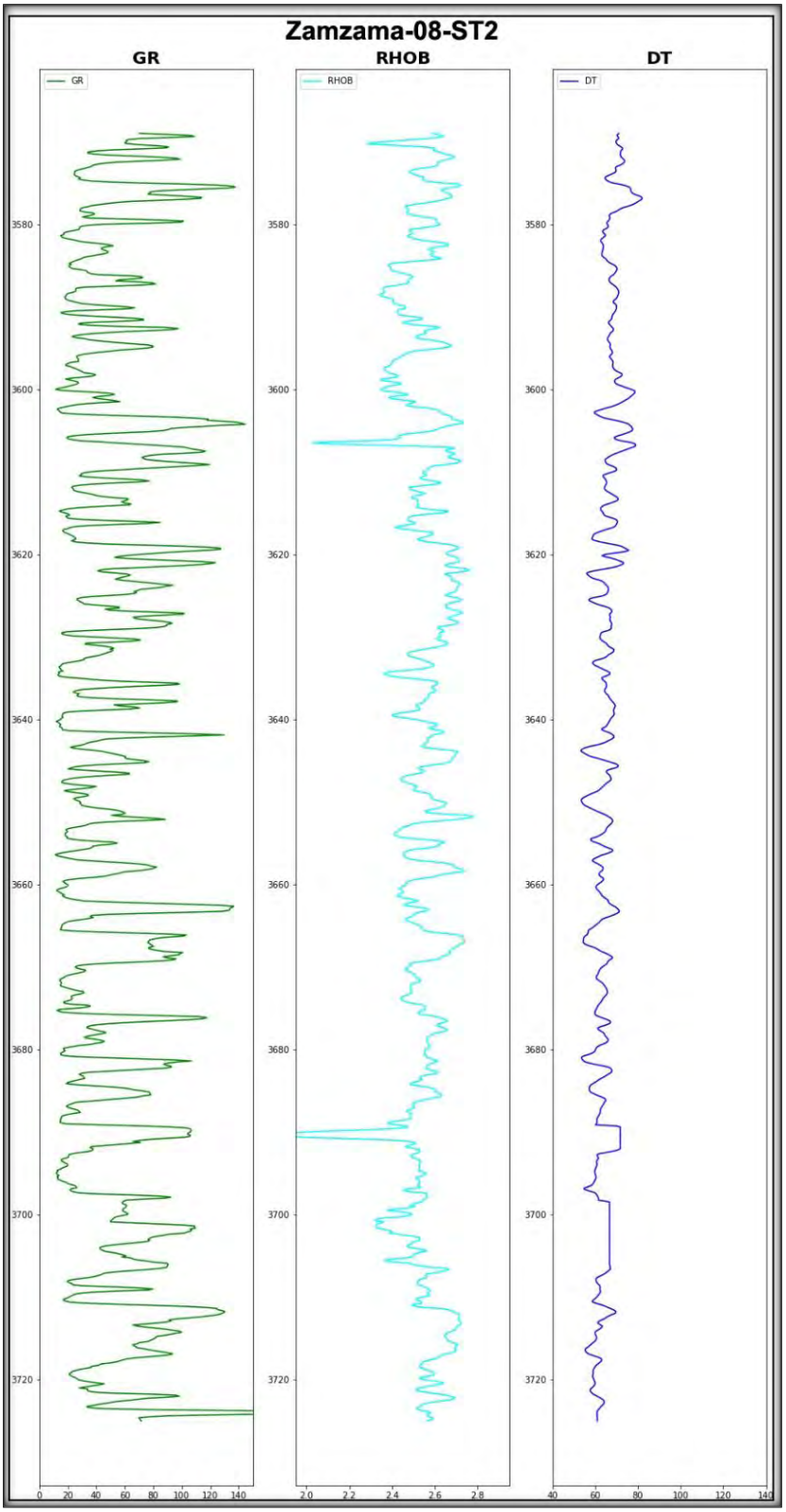


Figure 4.8 Input Dataset of machine learning for Zamzama- 08-ST2 used in testing for the prediction of VCL, PHIE, SW

4.3 Heat Map and Boxplots of Data

Heat maps and Boxplots are used to visualize the graphical display of data by colors. It is used to understand complex data at glance. In this research we are mainly working on GR, RHOB, DT used as input data while VCL, PHIE and SW curves are used as training and testing data. So make heat maps and boxplots of these curves for the understanding of data. In these boxplots there are some outliers but due to geological features we cannot remove the outliers. Using these plots we analyse the minimum, maximum and standard deviation of the datasets (Figure 4.9 to 4.13 for Zamzama 02, 03, 05, 06, 08-st2).

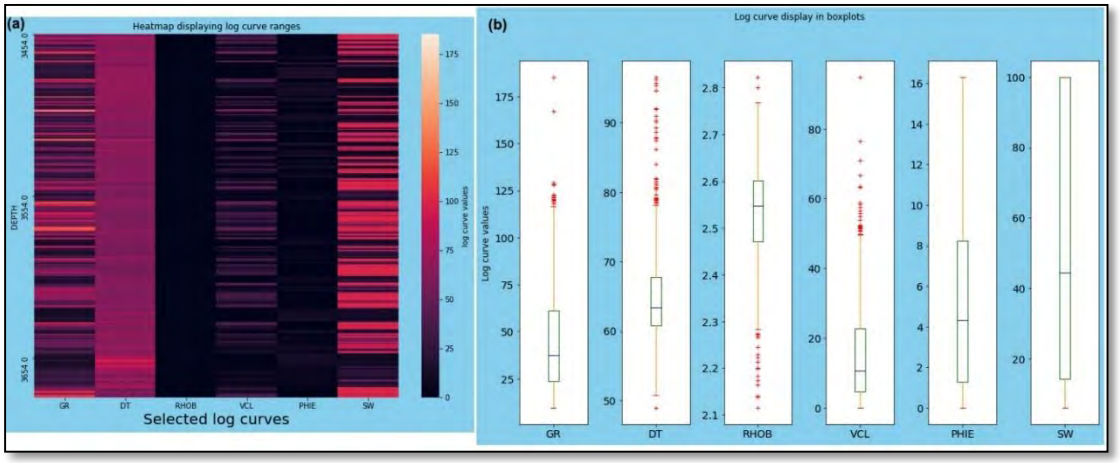


Figure 4.9 a) Heatmap of Zamzama-02 show values of GR, DT, RHOB, VCL, PHIE and SW using colors b) Boxplot of Zamzam-02 show different quantiles i.e. minimum, maximum and standard deviation of GR, DT, RHOB, VCL, PHIE and SW curves.

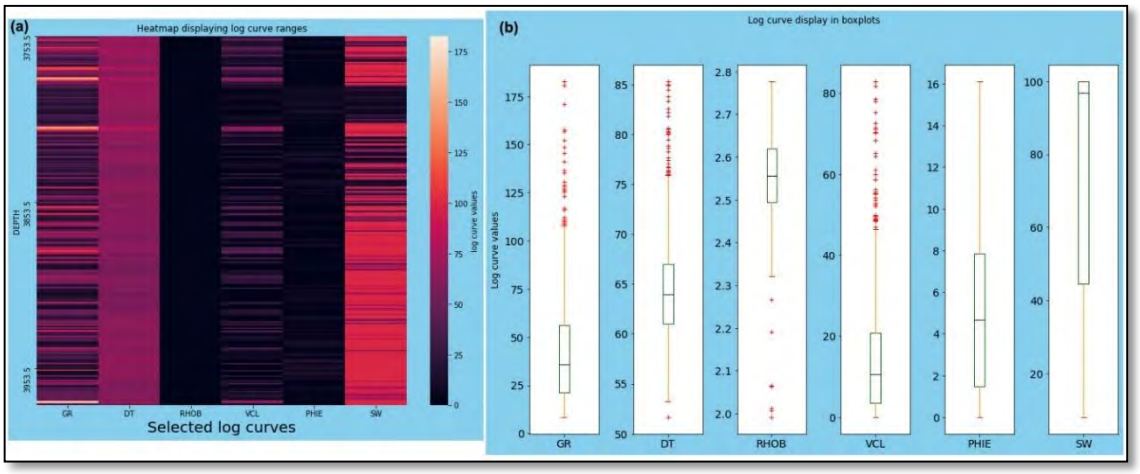


Figure 4.10 a) Heatmap of Zamzama-03 show values of GR, DT, RHOB, VCL, PHIE and SW using colors b) Boxplot of Zamzama-03 show different quantiles i.e. minimum, maximum and standard deviation of GR, DT, RHOB, VCL, PHIE and SW curves

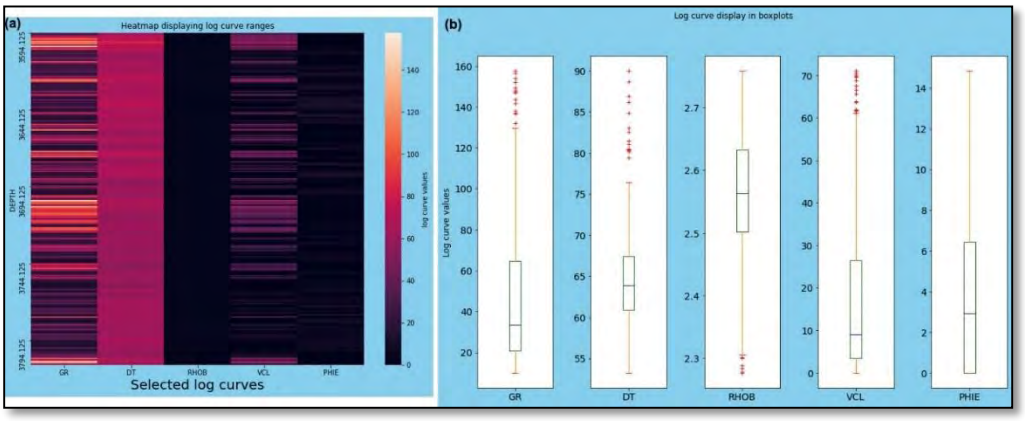


Figure 4.11 a) Heatmap of Zamzama-05 show values of GR, DT, RHOB, VCL, PHIE and SW using colors b) Boxplot of Zamzama -05 show different quantiles i.e. minimum, maximum and standard deviation of GR, DT, RHOB, VCL, PHIE and SW curves

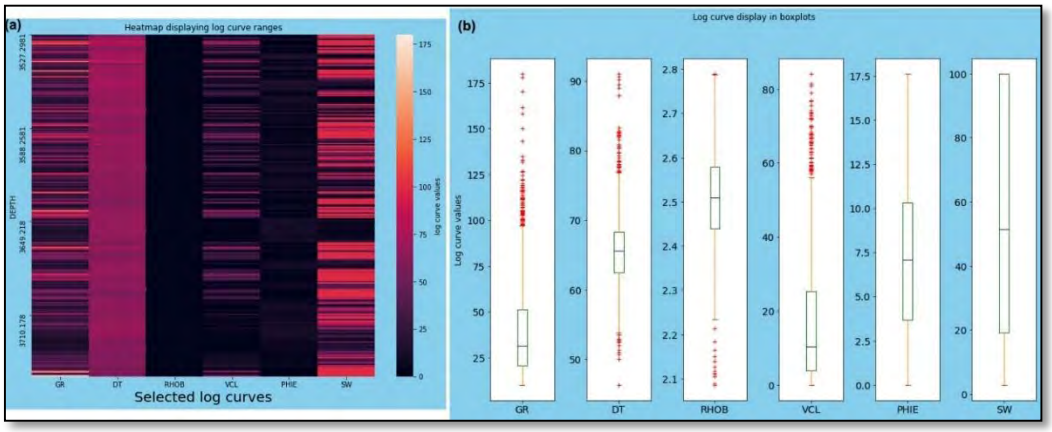


Figure 3 a) Heatmap of Zamzama -06 show values of GR, DT, RHOB, VCL, PHIE and SW using colors b) Boxplot of Zamzama -06 show different quantiles i.e. minimum, maximum and standard deviation of GR, DT, RHOB, VCL, PHIE and SW curves

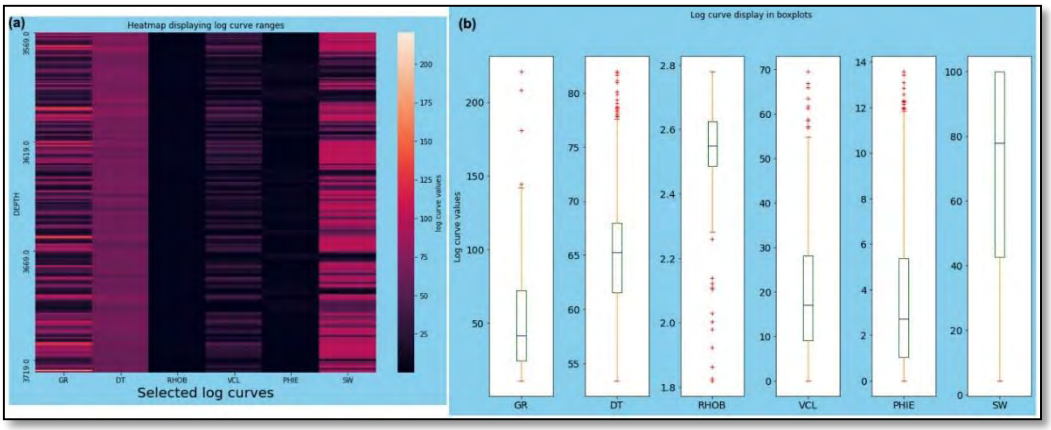


Figure 4.13 a) Heatmap of Zamzama -08-st2 show values of GR, DT, RHOB, VCL, PHIE and SW using colors b) Boxplot of Zamzama -08-st2 show different quantiles i.e. minimum, maximum and standard deviation of GR, DT, RHOB, VCL, PHIE and SW curves

4.4 Prediction of Petro-Elastic Properties at Wells Using Training Dataset

Different machine learning algorithms such as Random Forest, Decision Tree, SVR, XG boost are used to predict petro-elastic properties such as shear wave (S-wave), volume of clay (VCAL), Effective porosity (PHIE) and saturation of water (SW) at ZZ_05, ZZ_06 and ZZ_08 well by using ZZ-02 and ZZ-03 well as a training dataset. The R2 Score (%) of each algorithm is shown in tables.

4.4.1 Prediction of S-wave at Zamzama-05

By adopting different machine learning algorithms their correlation values between measured and predicted S-wave are compared for the selection of best method (Figure .14). The correlation scores i.e. RF=83, DTR=65, SVR=50, XG Boost=78 show that RF is best method for the prediction of S-wave as shown in table 2

Table 2 R2 Score (%) for prediction S-wave at Zamzama-05 for Random Forest (RF), Decision Tree (DTR), Support Vector Machine (SVM) and Xtreme Gradient Boost (Xgboost)

Algorithm	**RF**	**DTR**	**SVM**	**XGboost**
R2 Score (%)	83	65	50	78

.

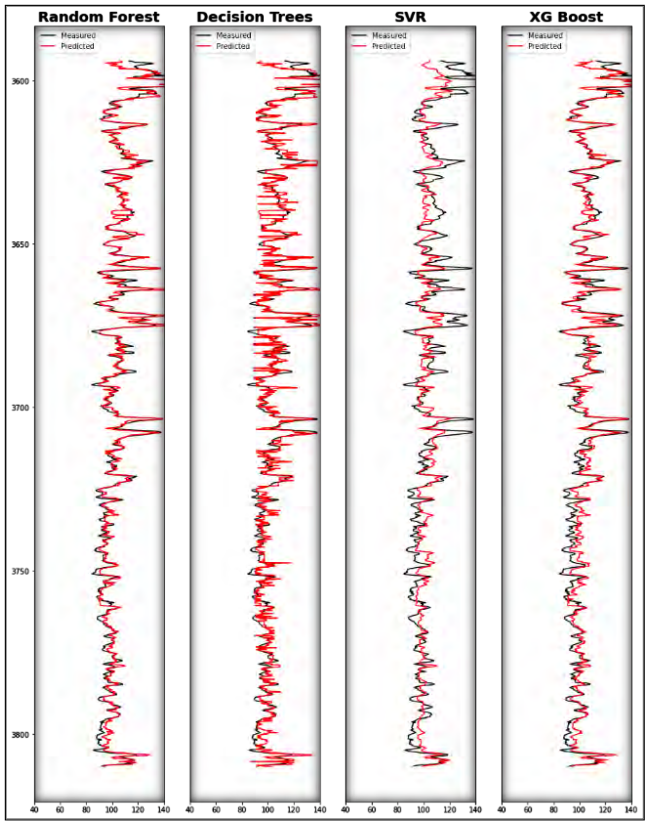


Figure 4.14 Prediction of S-wave at Zamzama -05 using Random Forest, Decision Tree, Support Vector Regression, Extreme Gradient Boost

4.4.2 Prediction of S-wave at Zamzama-06

In ZZ-06 well S-wave is missing so using all machine algorithms predict S-wave. RF and XG boost show excellent results (Figure 4.15).

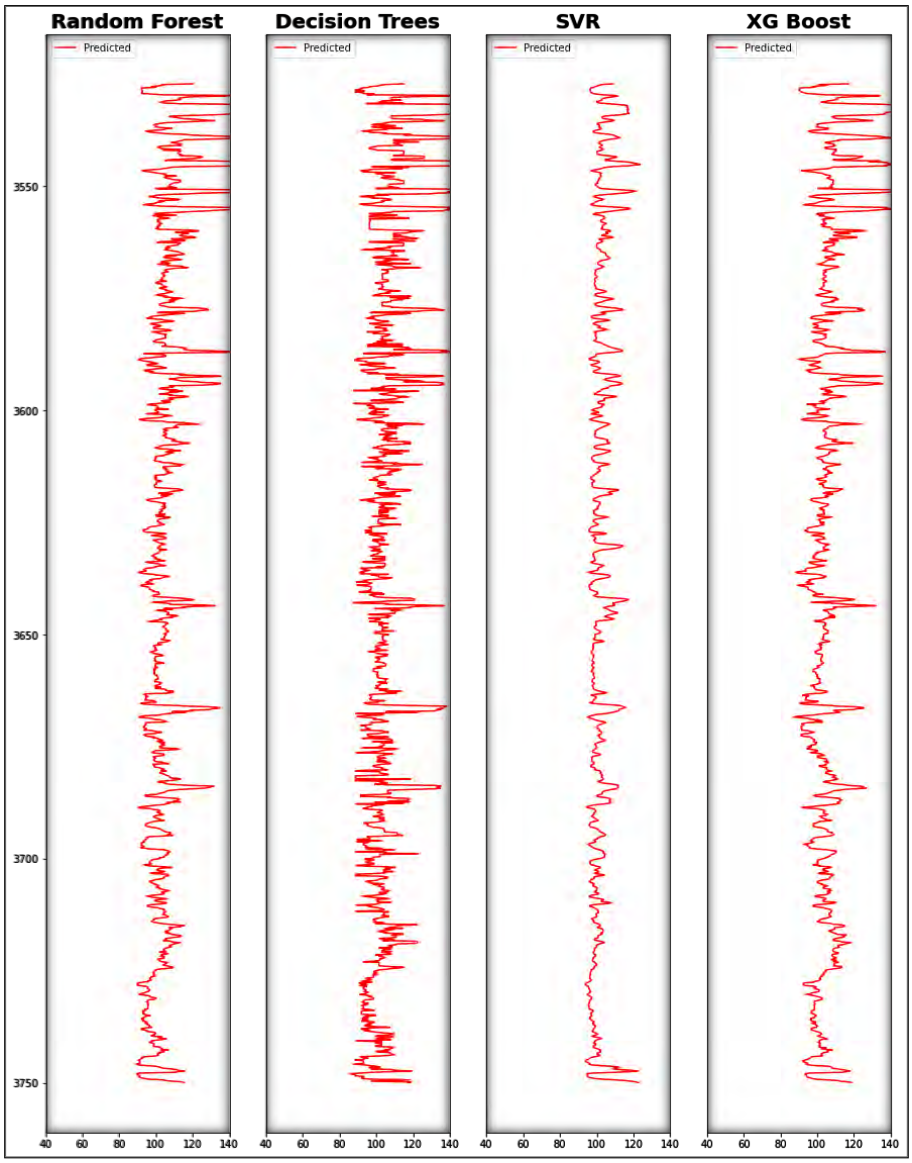


Figure 4.15 Prediction of S wave at Zamzama-06 using Random Forest, Decision Tree, Support Vector Regression, Extreme Gradient Boosting

4.4.3 Prediction of VCL at Zamzama-05

By adopting different machine learning algorithms their correlation values between measured and predicted VCL are compared for the selection of best method (Figure.4.16). The correlation scores i.e. RF=90, DTR=76, SVR=-3.1, XG Boost=89 show that RF is best method for the prediction of VCL (table 3).

Table 3 R2 Score (%) for prediction VCL at Zamzama-05 for Random Forest (RF), Decision Tree (DTR), Support Vector Machine (SVM) and Xtreme Gradient Boost (Xgboost)

Algorithm	**RF**	**DTR**	**SVM**	**XGboost**
R2 Score (%)	90	76	-3.1	89

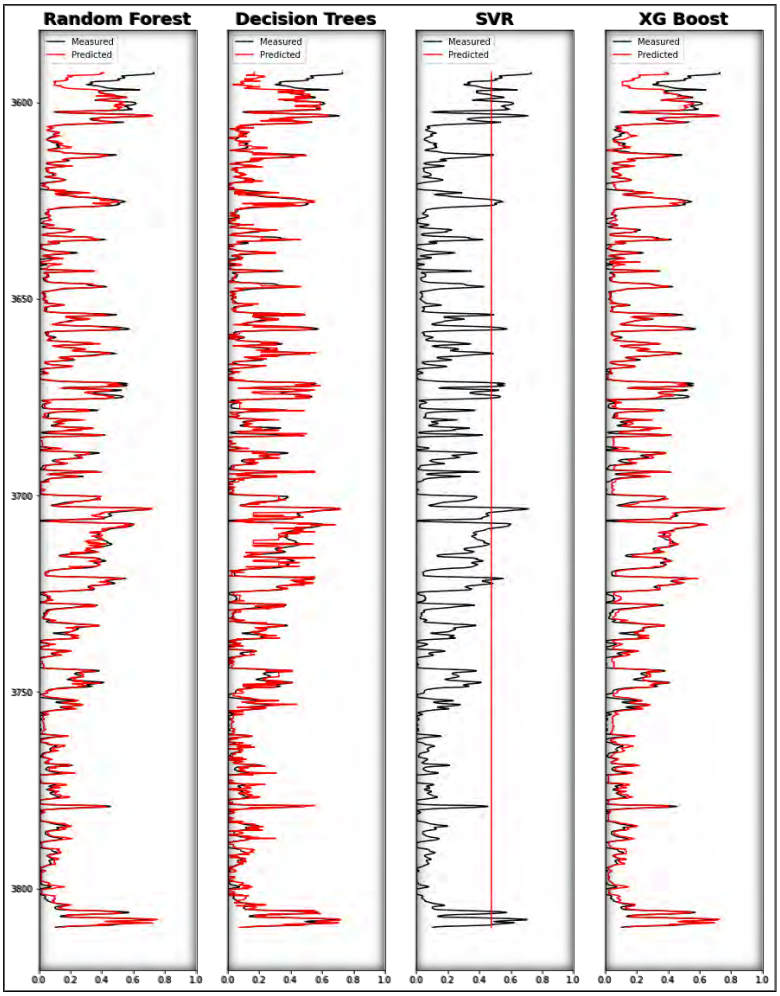


Figure 4.16 Prediction of VCL at Zamzama-05 using Random Forest, Decision Tree, Support Vector Regression, Extreme Gradient Boosting

4.4.4 Prediction of VCL at Zamzama-06

By adopting different machine learning algorithms their correlation values between measured and predicted VCL are compared for the selection of best method (4.17). The correlation scores i.e. RF=82, DTR=65, SVR=-2.9, XG Boost=80 show that RF is best method for the prediction of VCL (table 4).

Table 4 R2 Score (%) for prediction S-wave at Zamzama-05 for Random Forest (RF), Decision Tree (DTR), Support Vector Machine (SVM) and Xtreme Gradient Boost (Xgboost)

Algorithm	**RF**	**DTR**	**SVM**	**XGboost**
R2 Score (%)	82	65	-2.9	80

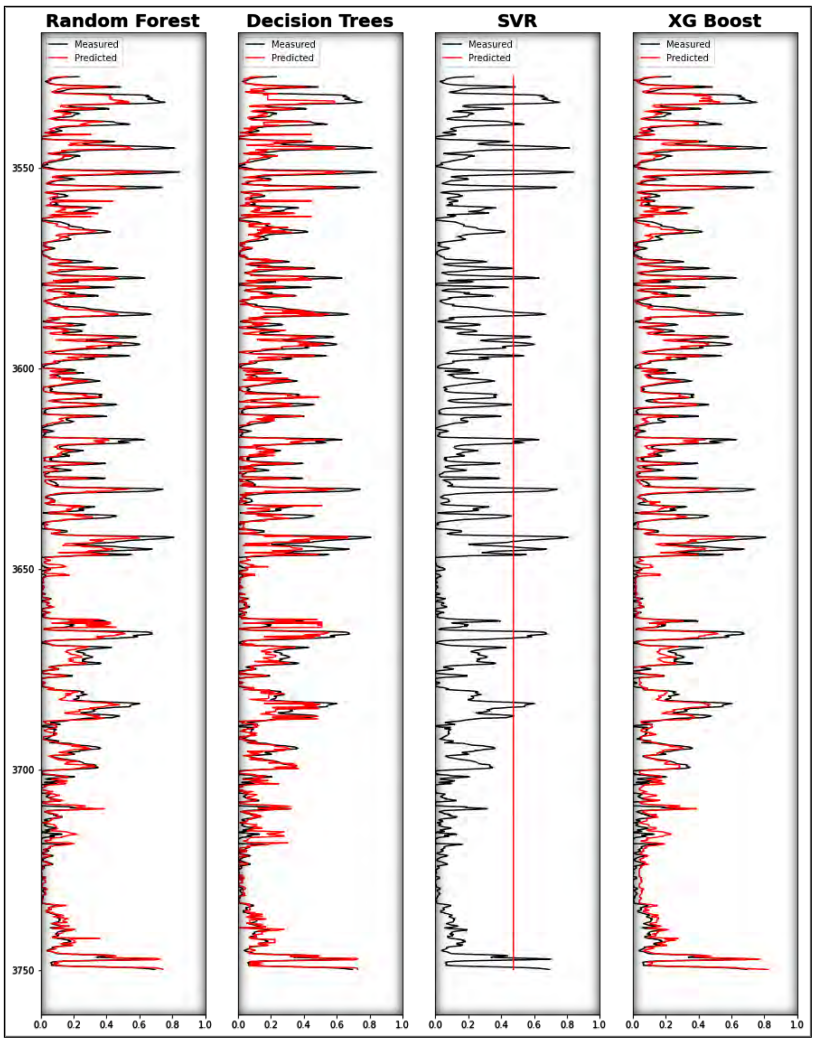


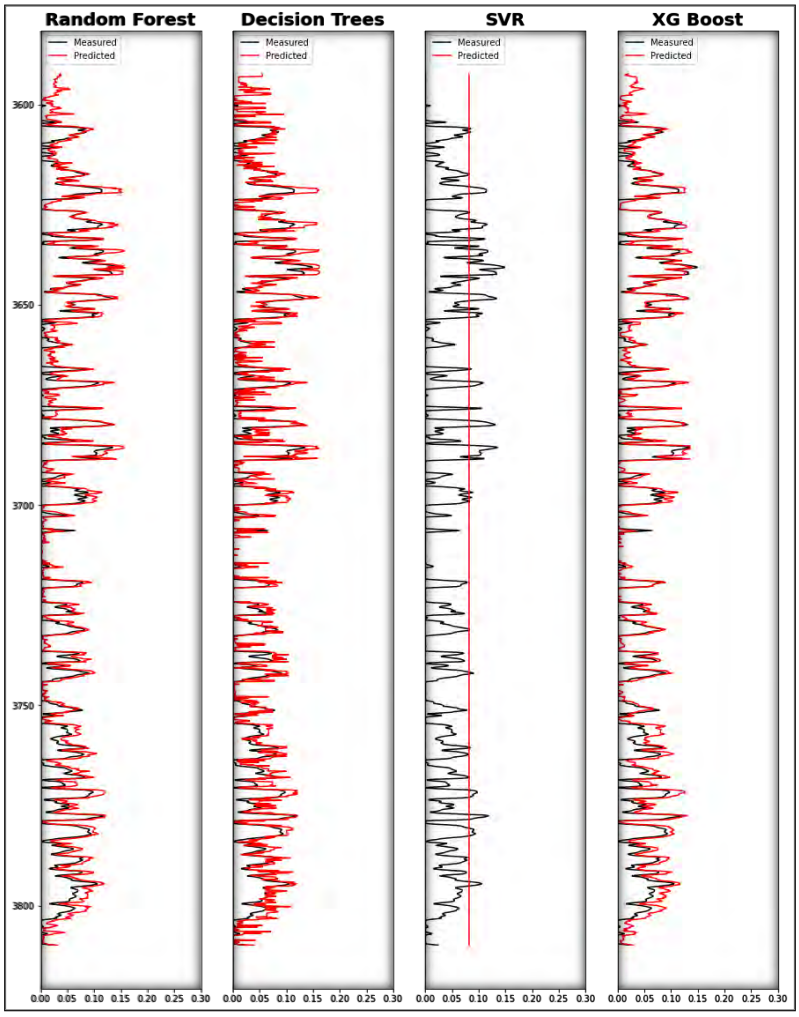
Figure 4.17 Prediction of VCL at Zamzama-06 using Random Forest, Decision Tree, Support Vector Regression, Extreme Gradient Boosting

4.4.5 Prediction of PHIE at Zamzama-05

By adopting different machine learning algorithms their correlation values between measured and predicted PHIE are compared for the selection of best method (Figure 4.18). The correlation scores i.e. RF=81.5, DTR=64, SVR=-1.49, XG Boost=77.7 show that RF is best method for the prediction of PHIE.

Table 5 R2 Score (%) for prediction of PHIE at Zamzama-05 for Random Forest (RF), Decision Tree (DTR), Support Vector Machine (SVM) and Xtreme Gradient Boost (Xgboost)

Algorithm	**RF**	**DTR**	**SVM**	**XGboost**
R2 Score (%)	81.5	64	-1.49	77.7



Figureure 4.18 Prediction of PHIE at Zamzama-05 using Random Forest, Decision Tree, Support Vector Regression, Extreme Gradient Boosting

4.4.6 Prediction of PHIE at Zamzama-06

By adopting different machine learning algorithms their correlation values between measured and predicted PHIE are compared for the selection of best method (4.20). The correlation scores i.e. RF=74.7, DTR=60.3, SVR=-9.2, XG Boost=70 show that RF is best method for the prediction of PHIE. (Table 6)

Table 6 R2 Score (%) for prediction of PHIE at Zamzama-06 for Random Forest (RF), Decision Tree (DTR), Support Vector Machine (SVM) and Xtreme Gradient Boost (Xgboost)

Algorithm	**RF**	**DTR**	**SVM**	**XGboost**
R2 Score (%)	74.7	60.3	-0.092	70

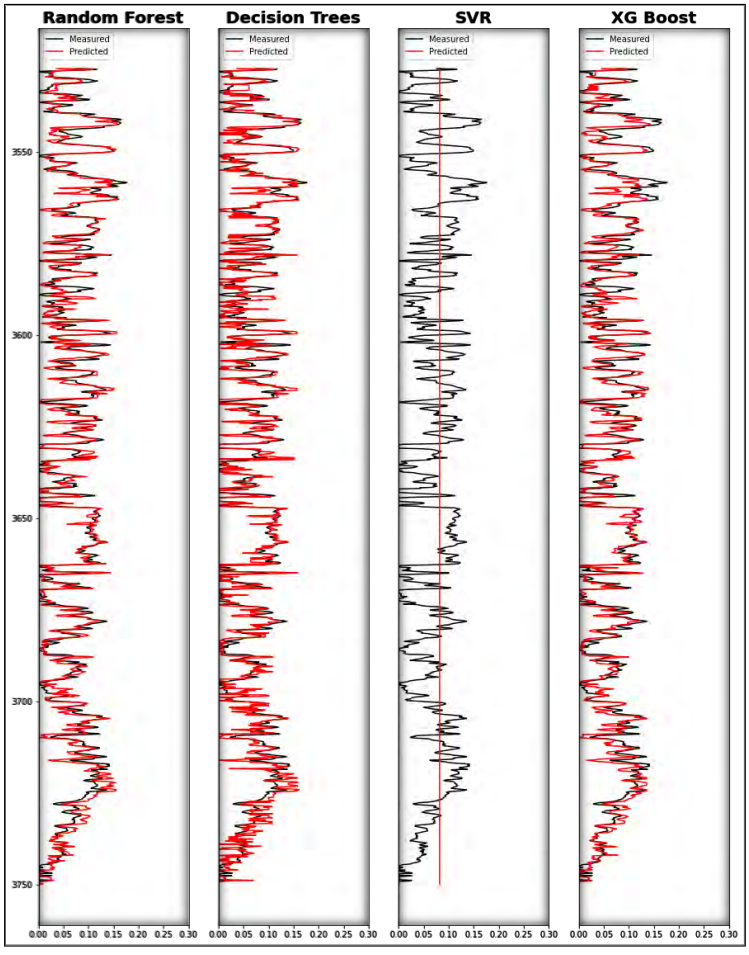


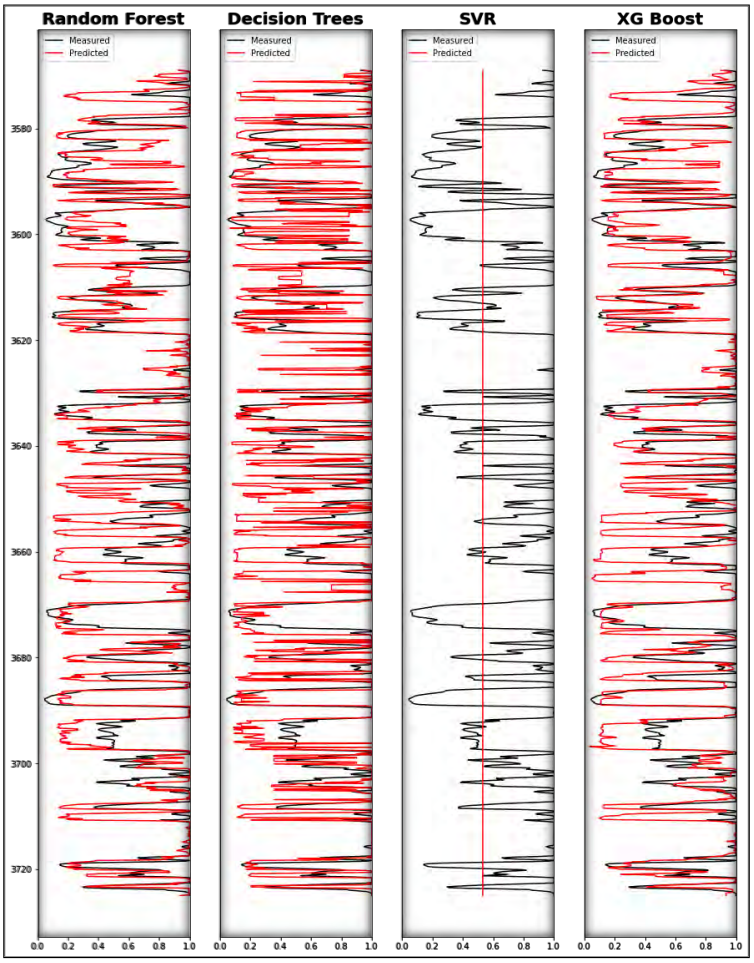
Figure 4.19 Prediction of PHIE at Zamzama-06 using Random Forest, Decision Tree, Support Vector Regression, Extreme Gradient Boosting

4.4.7 Prediction of SW at Zamzama-08-st2

By adopting different machine learning algorithms their correlation values between measured and predicted SW are compared for the selection of best method (4.20). The correlation scores i.e. RF=28, DTR=-6, SVR=-25, XG Boost=23 show that RF is best method for the prediction of SW (table 7).

Table 7 R2 Score (%) for prediction of SW at Zamzama-08-st2 for Random Forest (RF), Decision Tree (DTR), Support Vector Machine (SVM) and Xtreme Gradient Boost (Xgboost)

Algorithm	**RF**	**DTR**	**SVM**	**XGboost**
R2 Score (%)	28	-6	-25	23



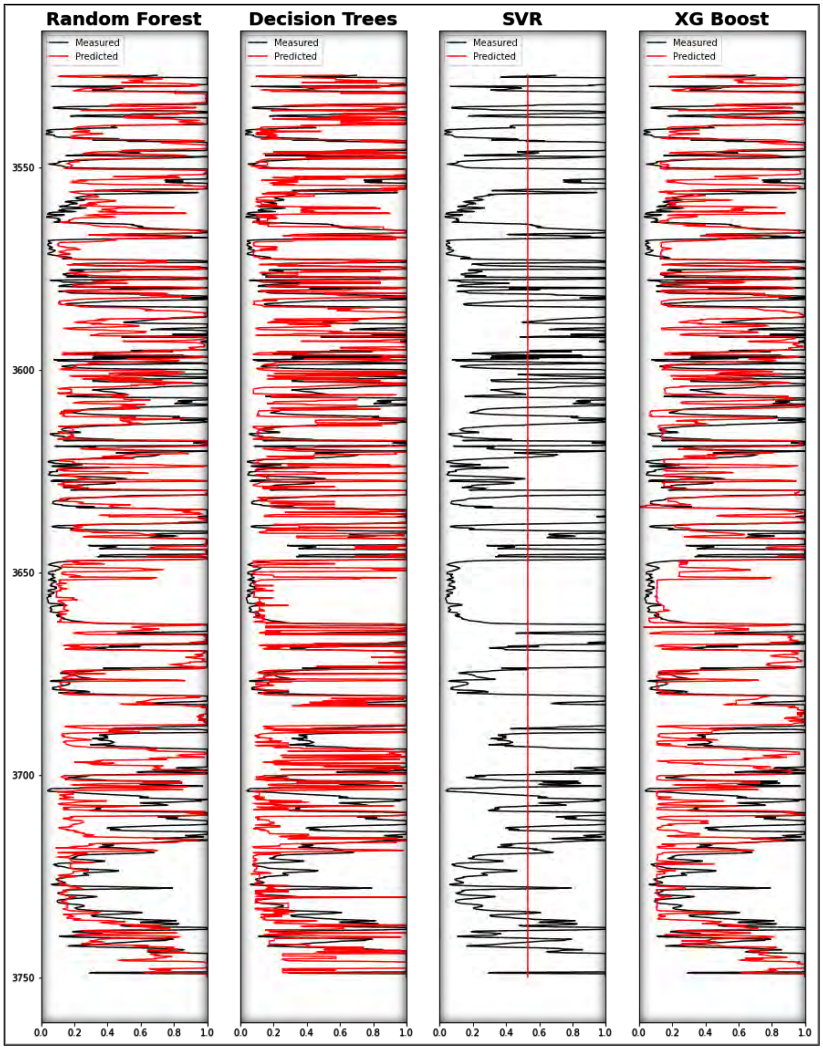
Figureure 4.20 Prediction of SW at Zamzama-08-st2 using Random Forest, Decision Tree, Support Vector Regression, Extreme Gradient Boosting

4.4.8 Prediction of SW at Zamzama-06

By adopting different machine learning algorithms their correlation values between measured and predicted SW are compared for the selection of best method (4.21). The correlation scores i.e. RF=42, DTR=14, SVR=-4, XG Boost=37.3 show that RF is best method for the prediction of SW(table 8).

Table 8 R2 Score (%) for prediction of SW at Zamzama-06 for Random Forest (RF), Decision Tree (DTR), Support Vector Machine (SVM) and Xtreme Gradient Boost (Xgboost)

Algorithm	**RF**	**DTR**	**SVM**	**XGboost**
R2 Score (%)	42	14	-4	37.3



Figureure 4.21 Prediction of SW at Zamzama-06 using Random Forest, Decision Tree, Support Vector Regression, Extreme Gradient Boosting

4.5 Property Modelling Using Machine Learning

Prediction of petrophysical properties at wells indicates that Random Forest Algorithm is best for the property modelling throughout the seismic cube. For this purpose, amplitude of seismic are used as input for the prediction of petrophysical properties (VCL, PHIE, SW) by training of Zamzama-03 and Zamzama-02.

4.5.1 Section View of VCL

A two-dimensional perspective view of petrophysical properties that are oriented from SE-NW direction of Zamzama gas field are shown in Figure 4.22. The section shows VCL ranges from 20-30%. Most of the wells have been drilled around at low VCL values and show good matched at blind wells like Zamzama-05, Zamzama -06 and Zamzama -08-st2. Yellow color show low values and green color show high values.

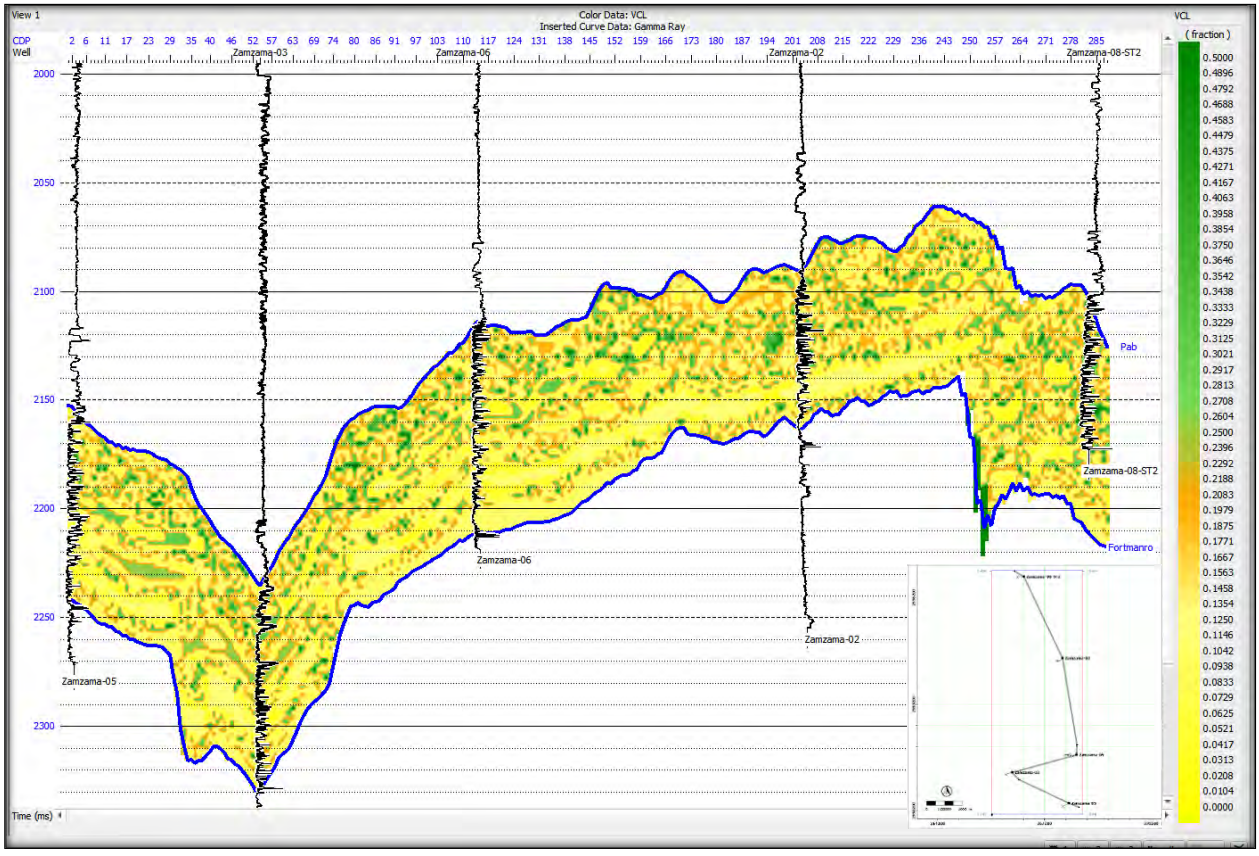


Figure 4.22 Section view of VCL by passing arbitrary line through all wells with display Gamma Ray log

4.5.2 Section View of PHIE

A two-dimensional perspective view of petrophysical properties that are oriented from SE-NW direction of Zamzama gas field are shown in Figure 4.23. The section shows PHIE ranges from 8-10%. Most of the wells have been drilled around at good PHIE values and show good matched at blind wells like Zamzama-05, Zamzama -06 and Zamzama -08-st2. Red color show high values of PHIE.

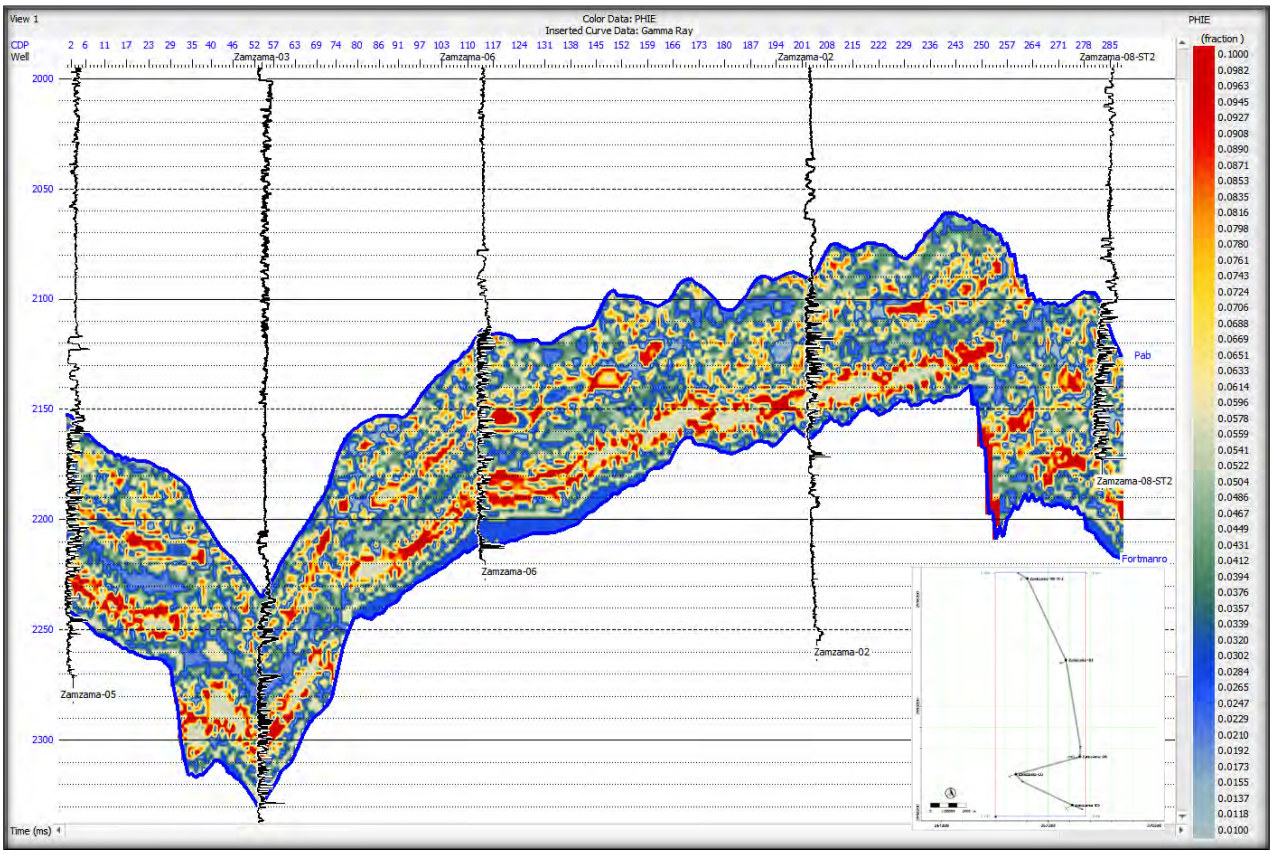
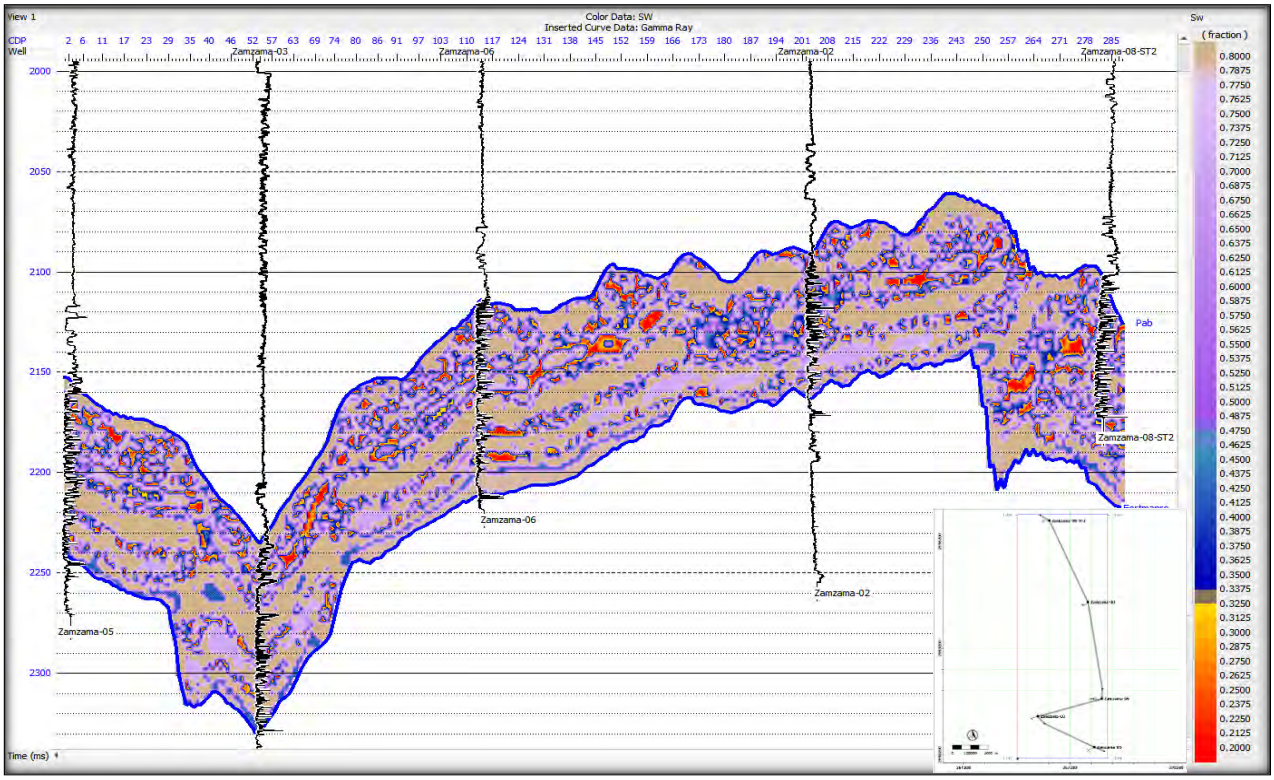


Figure 4.23 Section view of PHIE by passing arbitrary line through all wells with display Gamma Ray log

4.5.3 Section View of SW

A two-dimensional perspective view of petrophysical properties that are oriented from SE-NW direction of Zamzama gas field are shown in Figure 4.24. The section shows SW ranges from 40-

50%. Overall SW results are moderate because training at Zamzama-03 and Zamzama-08 is moderate and correlation is not good.



Figureure 4.24 Section view of SW by passing arbitrary line through all wells with display Gamma Ray log

4.6 Facies Modelling using K-Clusters

Facies modeling are done by using K-cluster Mean techniques. Three volumes cube i.e. VCL, PHIE, SW are used as an input for K-Clustering for distribution of facies on the basis of prediction data of VCL, PHIE and SW. These results are matched with well litho-Facies.

4.6.1 Clustering

Data of VCL, PHIE and SW are clustered into 3 cluster sets. Red color indicates Gas, Green color show shale and blue color show wet sand in Figure 4.25.

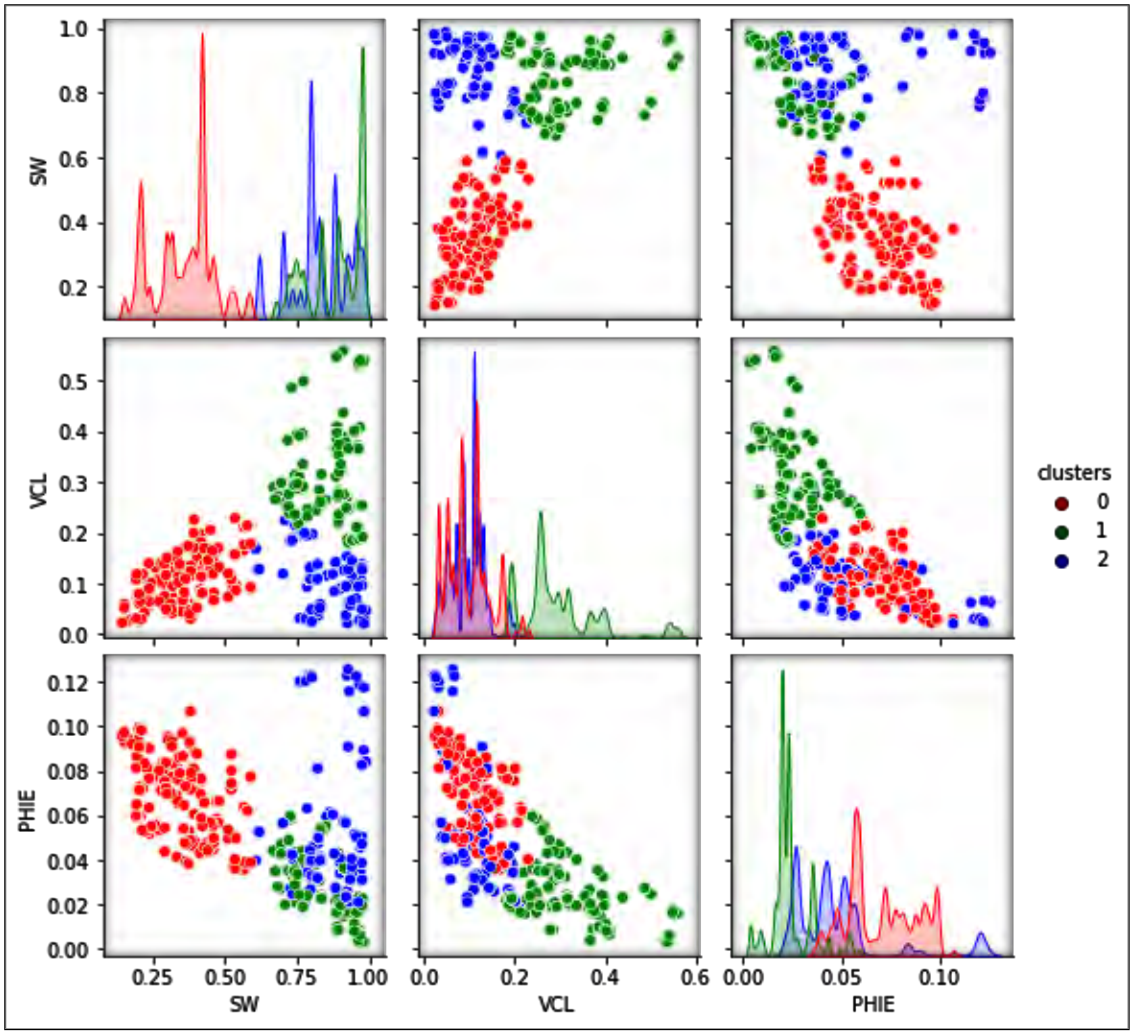


Figure 4.25 clustering of VCL, PHIE and SW cube dataset predicted by Random Forest Machine Learning Algorithms

4.6.2 QC at Well

The homogeneity score between litho-facies Curve (LFC) and K-cluster facies at Zamzama-03 is 78% while homogeneity score between litho-facies Curve (LFC) and K-cluster facies at Zamzama -02 is 82%. This is the excellent Matched between LFC and facies (Figure 4.26). In first track VCL curve run, in second track PHIE curve while in third track SW curve run. And the last two track show the comparison of facies marked by well curves and K-clustering technique.

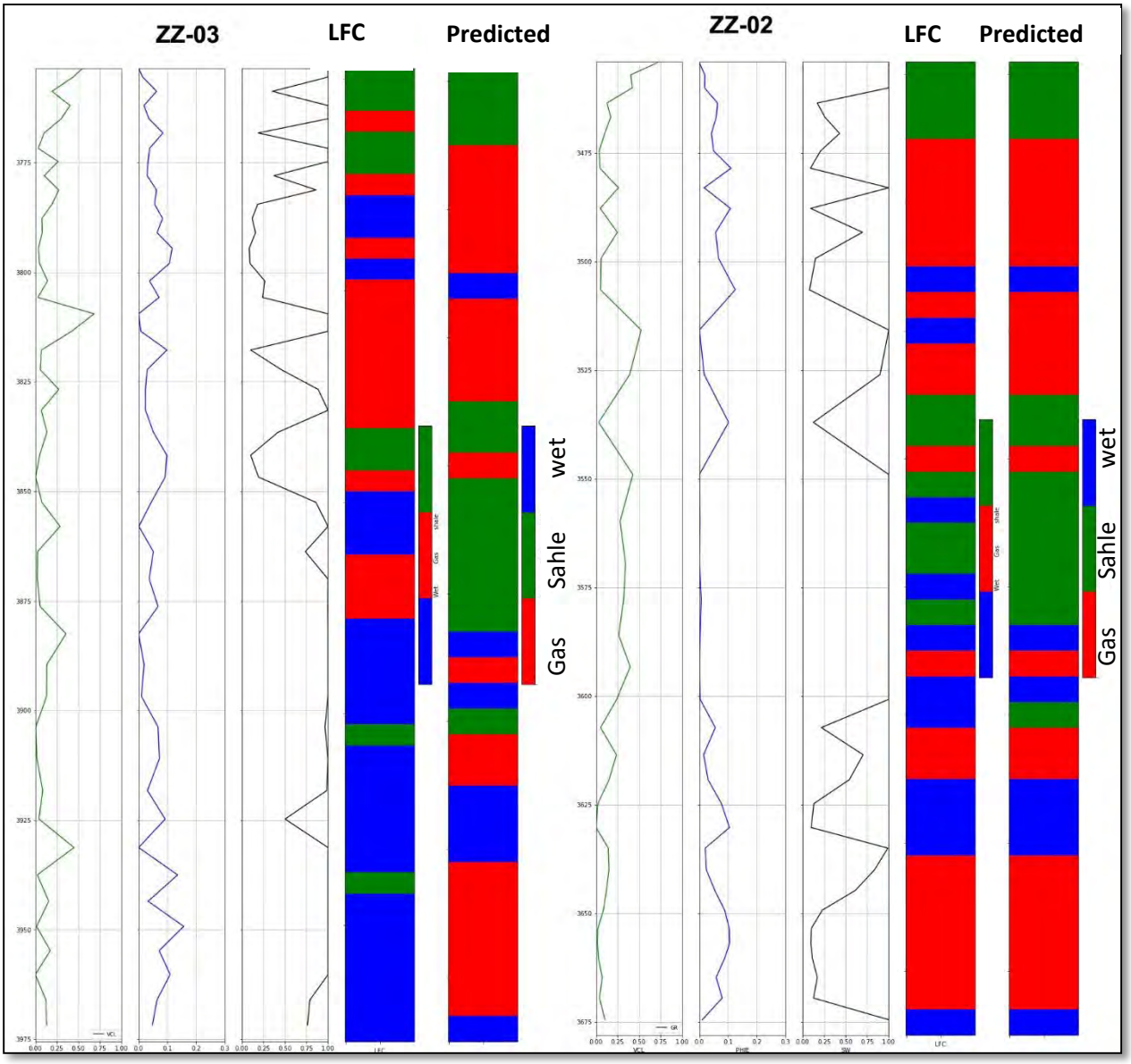


Figure 4.26 QC plot between Litho-Facies curves and K-clustering facies at well Zamzama-03 and Zamzama-02. Red color show Gas sand, green color show shale and blue color show Wet sand.

The distribution of facies calculated by K-clustering technique at inline where wells was drilled. Zamzama -03 is drilled at inline=482 and xline=146, Zamzama -02 is drilled at inline=535 and xline=193, Zamzama -05 is drilled at inline=398 and xline=199, Zamzama -06 is drilled at inline=445 and xline208 and Zamzama -08St2 is drilled at inline=613 and xline=156. The facies distribution is shown in Figure 4.27 for Zamzama 03 and 02, for Zamzama 05 and 06 in Figure. 4.28 Whereas for Zamzama-08-st2 in Figure 4.29

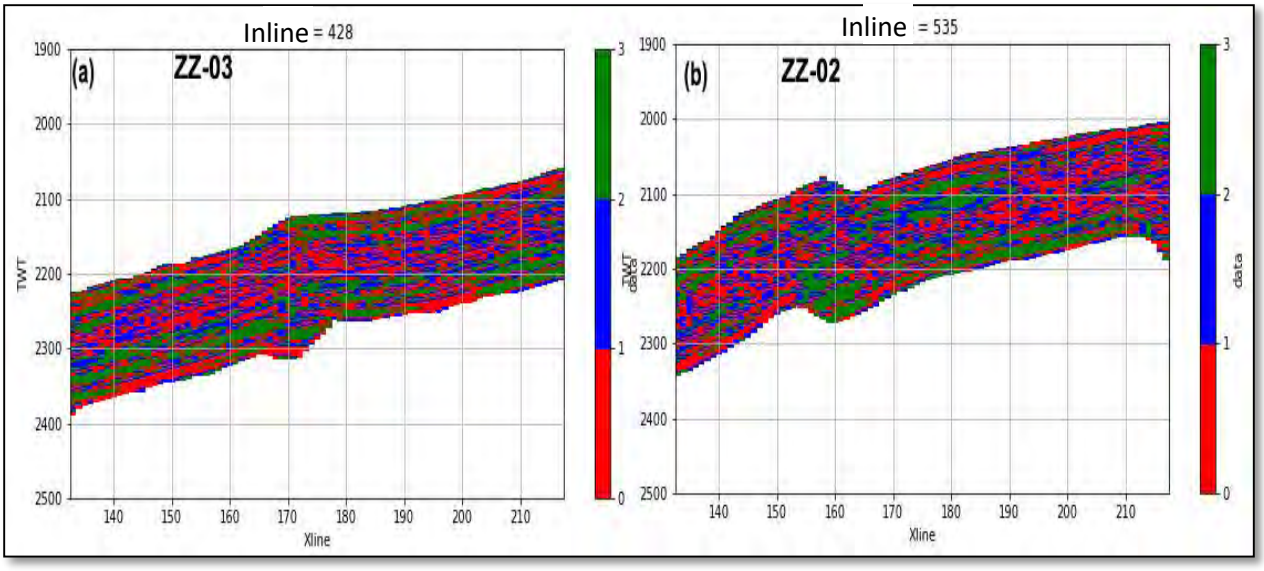


Figure 4.27 a) Facies distribution calculated by K-cluster technique at Zamzama -03 b) Facies distribution calculated by K-cluster technique at Zamzama -02. Red color show Gas sand, green color show shale and blue color show Wet sand.

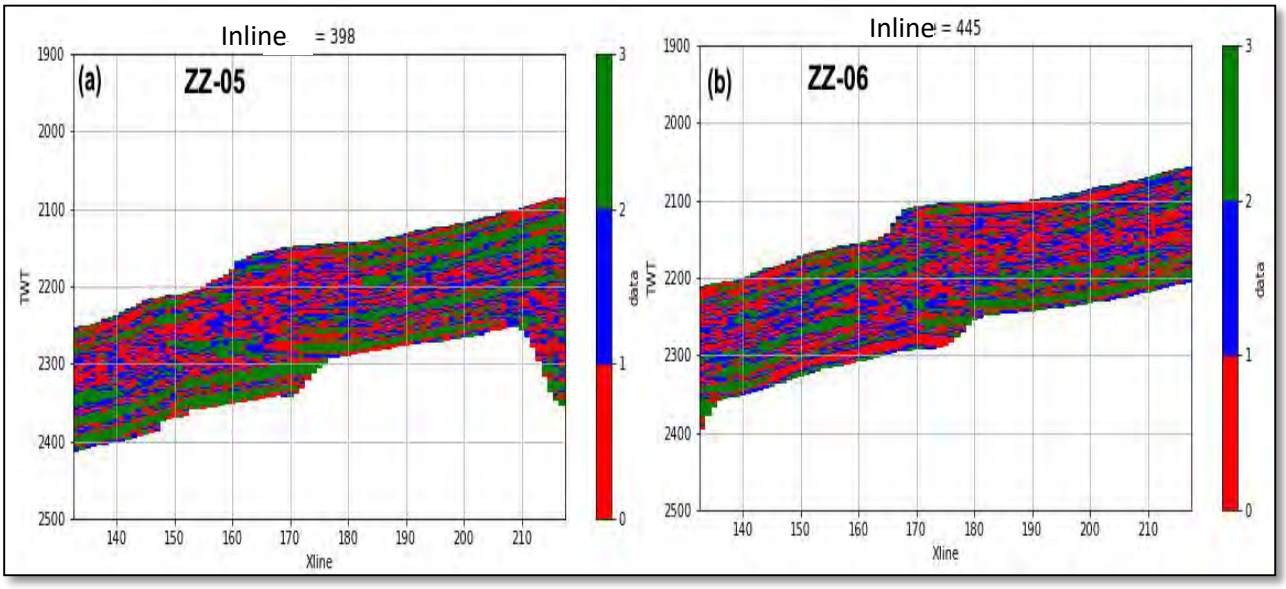


Figure 4 a) Facies distribution calculated by K-cluster technique at Zamzama -05 b) Facies distribution calculated by K-cluster technique at Zamzama -06. Red color show Gas sand, green color show shale and blue color show Wet sand.

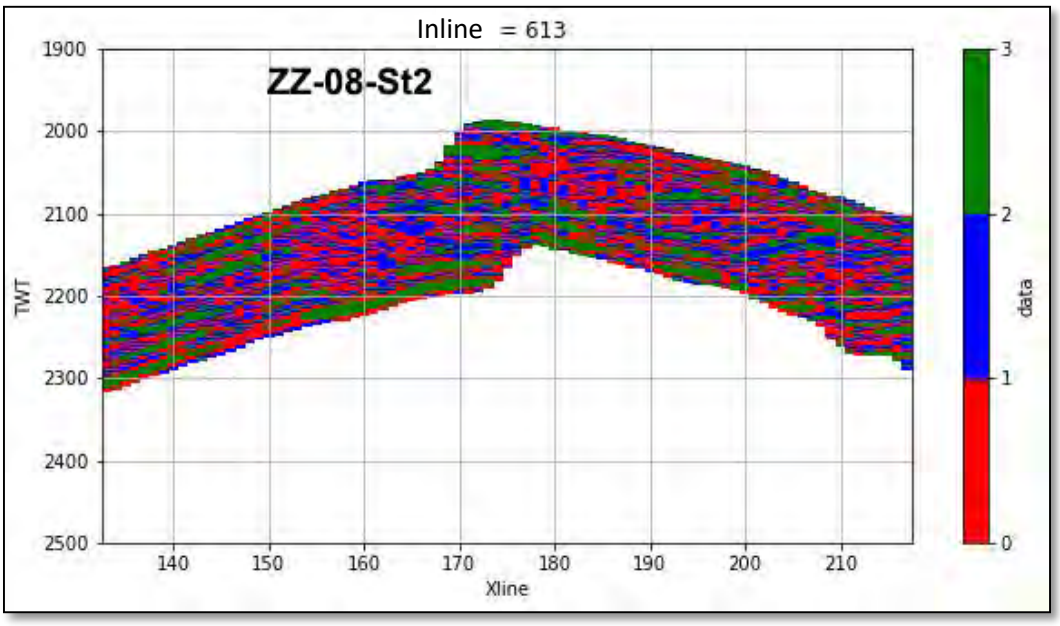


Figure 4.29 Facies distribution calculated by K-cluster technique at Zamzama -08St2. Red color show Gas sand, green color show shale and blue color show Wet sand.

4.7 Conclusion and Discussion

A comprehensive reservoir characterization study was carried out on the Zamzama gas field with the help of advanced machine learning algorithms. A total of five (5) wells were used in this study namely, Zamzama-02, Zamzama-03, Zamzama-05, Zamzama-06, and Zamzama-08-ST2. Initially detailed petrophysical interpretation was carried out. These results were discussed in detail which were also aligned with the actual production details.

The essential log required for reservoir characterization is S-wave, which was missing from ZZ-06 well. This was predicted using the Random Forest machine learning algorithm. Initially the same was predicted for ZZ-05 and matched with the measured S-wave log available at reservoir level. Based on the excellent match obtained, S-wave was predicted for ZZ-06.

In order to predict the key petrophysical properties across the cube the machine learning random forest algorithm was trained on two selected wells of ZZ-02 & ZZ-03, which provided us a good match of ~80% and above.

Taking these petrophysical volumes as input to K-Mean clustering algorithm, facies modelling was carried out. This phase not only provided us with the essential QC of the petrophysical properties, but also help identify facies which were in direct match with the well lithofacies curve.

Therefore, it is concluded that the implementation of advanced machine learning algorithms not only helped us save a lot of computation time and workflow, but also help with increased accuracy and efficiency. This is indeed the future of the oil and gas exploration in Pakistan

5 REFERENCES

- Abbasi, S. A., Asim, S., Solangi, S. H., & Khan, F. (2016). Study of fault conFigureuration related mysteries through multi seismic attribute analysis technique in Zamzama gas field area, southern Indus Basin, Pakistan. Geodesy and Geodynamics, 7(2), 132-142.
- Ahmad, A., & Dey, L. (2007). A k-mean clustering algorithm for mixed numeric and categorical data. Data & Knowledge Engineering, 63(2), 503-527.
- Alpaydin Ethem. Introduction to Machine Learning, 2014. Introduction to Machine Learning (3rd Edition) p.3.
- Anifowose, F.A., Labadin, J., and Abdulraheem, A., 2017, Ensemble machine learning: An untapped modeling paradigm for petroleum reservoir characterization: Journal of petroleum science & engineering, v. 151, p. 480–487, doi:10.1016/j.petrol.2017.01.024.
- Awad, M., and R. Khanna, 2015, Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers: Apress, Berkeley, CA, XIX, 268 p
- Ayodele, T. O. (2010). Types of machine learning algorithms. New advances in machine learning, 3, 19-48.
- Camps-Valls, G., 2020, Advances in machine learning for modelling and understanding in earth sciences:
- Caté, A., Perozzi, L., Gloaguen, E., & Blouin, M. (2017). Machine learning as a tool for geologists. The Leading Edge, 36(3), 215-219.
- Chaki, S., 2015, Reservoir Characterization: A machine learning approach: arXiv [cs.CE], http://arxiv.org/abs/1506.05070.
- Chaki, S., Routray, A., and Mohanty, W.K., 2018, Well-log and seismic data integration for reservoir characterization: A signal processing and machine-learning perspective: IEEE signal processing magazine, v. 35, p. 72–81, doi:10.1109/msp.2017.2776602.
- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable Tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
- Costa, I. S. L., Tavares, F. M., & de Oliveira, J. K. M. (2019). Predictive lithological mapping through machine learning methods: a case study in the Cinzento Lineament, Carajás Province, Brazil. Journal of the Geological Survey of Brazil, 2(1), 26-36.
- Culverhouse, P. F., Williams, R., Reguera, B., Herry, V., & González-Gil, S. (2003). Do experts make mistakes? A comparison of human and machine indentification of dinoflagellates. Marine ecology progress series, 247, 17-25.

- Feng, R., Grana, D., and Balling, N., 2021, Imputation of missing well log data by random forest and its uncertainty analysis: Computers & geosciences, v. 152, p. 104763, doi:10.1016/j.cageo.2021.104763.
- Fridman, L., Brown, D. E., Glazer, M., Angell, W., Dodd, S., Jenik, B., ... & Reimer, B. (2019). MIT advanced vehicle technology study: Large-scale naturalistic driving study of driver behavior and interaction with automation. IEEE Access, 7, 102021-102038.
- Friedl, M.A., and Brodley, C.E., 1997, Decision Tree classification of land cover from remotely sensed data, Remote Sensing of Environment, Volume 61, Issue 3, Pages 399-409, https://doi.org/10.1016/S0034-4257(97)00049-7.
- Hulbert, C., Rouet-Leduc, B., Jolivet, R., & Johnson, P. A. (2020). An exponential build-up in seismic energy suggests a months-long nucleation of slow slip in Cascadia. Nature communications, 11(1), 1-8.
- Hussein, M., Stewart, R. R., Sacrey, D., Wu, J., & Athale, R. (2021). Unsupervised machine learning using 3D seismic data applied to reservoir evaluation and rock type identification. Interpretation, 9(2), T549-T568.
- Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020). Supervised machine learning: a brief primer. Behavior Therapy, 51(5), 675-687.
- Latifovic, R., Pouliot, D., & Campbell, J. (2018). Assessment of convolution neural networks for surficial geology mapping in the South Rae geological region, Northwest Territories, Canada. Remote sensing, 10(2), 307.
- Liaw, A., and Wiener, M., 2002, Classification and Regression by random Forest, R news, Vol.2/3.
- Marjanović, M., Kovačević, M., Bajat, B., & Voženílek, V. (2011). Landslide susceptibility assessment using SVM machine learning algorithm. Engineering Geology, 123(3), 225-234.
- Meier, U., Curtis, A., & Trampert, J. (2007). Fully nonlinear inversion of fundamental mode surface waves for a global crustal model. Geophysical Research Letters, 34(16).
- Merembayev, T., Yunussov, R., & Yedilkhan, A. (2018, November). Machine learning algorithms for classification geology data from well logging. In 2018 14th International Conference on Electronics Computer and Computation (ICECCO) (pp. 206-212). IEEE.
- Nikhil Buduma (2017). Fundamentals of deep learning: designing next generation machine intelligence algorithms p. 3.
- Nwachukwu, C., 2018, Machine learning solutions for reservoir characterization, management, and optimization (Doctoral dissertation):
- Pal, M., 2005, Random Forest classifier for remote sensing classification, International Journal of Remote Sensing, volume 26, no. 1, page 217 to 222, DOI: 10.1080/01431160412331269698.

- Palafox, L. F., Hamilton, C. W., Scheidt, S. P., & Alvarez, A. M. (2017). Automated detection of geological landforms on Mars using Convolutional Neural Networks. *Computers & geosciences*, *101*, 48-56.
- Qureshi, M.A., Ghazi, S., Riaz, M., and Ahmad, S., 2020, Geo-seismic model for petroleum plays an assessment of the Zamzama area, Southern Indus Basin, Pakistan: Journal of petroleum exploration and production technology, doi:10.1007/s13202-020-01044-7.
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. J. O. G. R. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression Trees and support Vector Machines. Ore Geology Reviews, 71, 804-818.
- Safavian, S. R., and Landgrebe, D., 1991, A survey of decision Tree classifier methodology, in IEEE Transactions on Systems, Man, and Cybernetics, vol. 21, no. 3, pp. 660-674, May-June1991.
- Thessen, A. (2016). Adoption of machine learning techniques in ecology and earth science. One Ecosystem, 1, e8621.
- Wang, P., Chen, X., Wang, B., Li, J., and Dai, H., 2020, An improved method for lithology identification based on a hidden Markov model and random forests: Geophysics, v. 85, p. IM27–IM36, doi:10.1190/geo2020-0108.1.