A MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE APPROACH TO ESTIMATE GEOMECHANICAL PARAMETERS FROM CORE SAMPLES

The thesis is submitted in partial fulfilment of the requirements of

Bachelor of Technology (Petroleum Engineering)

by

DHARTI PATEL	**17BPE024**
KRIPAL PATEL	**17BPE074**
JWALIT SOLANKI	**17BPE118D**
GAIRIK CHAKRABORTY	**17BPE032**
KAPIL SORATHIYA	**17BPE105**
PRADYUMNA LOKIN	**17BPE053**
RAVI RATHOD	**17BPE094**

Under the Guidance of

Dr Paul Naveen
Dr Jwnsagar Brahma

School of Petroleum Technology

Pandit Deendayal Energy University
Gandhinagar- 382007, Gujarat, India

JUNE 2021

APPROVAL SHEET

This thesis entitled **"A Machine Learning and Artificial Intelligence Approach to Estimate Geomechanical Parameters from core samples"** by **Dharti Patel** (17BPE024), **Kripal Patel** (17BPE074), **Jwalit Solanki** (17BPE118D), **Gairik Chakraborty** (17BPE032), **Kapil Sorathiya** (17BPE105), **Pradyumna Lokineni** (17BPE053) and **Ravi Rathod** (17BPE094) are recommended for the degree of Bachelor of Technology in Petroleum Engineering.

Examiners

_____

_____

_____

Supervisors

_____

_____

_____

Chairman

_____

Date:

Place:

DECLARATION

We, Dharti Patel, Kripal Patel, Jwalit Solanki, Gairik Chakraborty, Kapil Sorathiya, Pradyumna Lokineni, and Ravi Rathod, hereby declare that this written submission represents our ideas in our own words and where others idea or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all academic integrity principles and integrity and have not misrepresented or fabricated or falsified any ideas/data/fact source in my submission. We understand that any violation of the above will cause disciplinary action by the PANDIT DEENDAYAL ENERGY UNIVERSITY and evoke penal action from the source that has thus not been appropriately cited, which proper permission has not been taken needed.

_____

Dharti Patel (17BPE024)

_____

Kripal Patel (17BPE074)

_____

Jwalit Solanki (17BPE0118D)

_____

Gairik Charkraborty (17BPE032)

_____

Kapil Sorathiya (17BPE105)

_____

Pradyumna Lokineni (17BPE053)

_____

Ravi Rathod (17BPE094)

Place:

Date:

CERTIFICATE

This is to certify that the project entitled **"A Machine Learning and Artificial Intelligence Approach to Estimate Geomechanical Parameters from core samples"** is the bonafide carried out by Dharti Patel (17BPE024), Kripal Patel (17BPE074), Jwalit Solanki (17BPE118D), Gairik Chakraborty (17BPE032), Kapil Sorathiya (17BPE105), Pradyumna Lokineni (17BPE053) and Ravi Rathod (17BPE094) for the fulfilment of the requirement for the degree of Bachelor of Technology in Petroleum Engineering, School of Petroleum Technology, Pandit Deendayal Energy University, Gandhinagar during the year 2020-2021. This work was carried out under my guidance and supervision. No part of the dissertation has been submitted for the award of any degree or diploma elsewhere to the best of my knowledge.

_____

Signature of Mentor

(Dr. Paul Naveen)

_____

Signature of Co-Mentor

(Dr. Jwngsar Brahma)

Date:

Place:

ACKNOWLEDGEMENT

We would like to express our gratitude to our supervisors, Dr. Paul Naveen, Assistant Professor, SPT, PDEU and Dr. Jwngsar Brahma, Assistant Professor, SOT, PDEU for continuously guiding us through the project. We are thankful to Dr. Rakesh Kumar Vij, Director, SPT, PDEU, to work on this project. Also, we would like to thank the SPT Faculty, our classmates and our parents for their continuous and unwavering support.

Dharti Patel (17BPE024)

Kripal Patel (17BPE074)

Jwalit Solanki (17BPE118D)

Gairik Chakrborty (17BPE032)

Kapil Sorathiya (17BPE105)

Pradyumna Lokineni (17BPE053)

Ravi Rathod (17BPE094)

B.Tech 4th Year, Petroleum Engineering

School of Petroleum Technology

Pandit Deendayal Energy University,

Raisan, Gandhinagar-382007

MOTIVATION FOR PROJECT

Many industrial sectors have faced a severe meltdown due to the COVID-19 global pandemic. The Oil and Gas sector is no exception, and there has been a lot of uncertainty in 2020 on Oil and Gas, which can primarily be associated with the pandemic. The disruption in the global supply, crumpling of the transportation system, and the steep decrease in the hostile prices of the WTI crude due to volatility in demand.

The sector has faced many challenges, but there is a silver lining. The pandemic has led us to understand the importance of the term 'digitalisation' and its implementation on various domains which were not explored much in recent decades. Working from a remote area for an Oil and Gas company, automation is a demanding technology that can significantly improve operations and cut costs in the upstream Oil and Gas industry. The concept of Big data and Artificial intelligence is no stranger to this sector, but this pandemic has led to its quick and rapid adoption and innovation in these new cutting-edge technologies. Digital transformation in the industry can create more revenues with productivity improvements.

Geomechanical parameters are crucial information while designing many petroleum operations like hydraulic fracturing, wellbore stability model, prediction of in-situ stress, and more. The direct measurements of these geomechanical parameters are sometimes not viable because of the high cost of tests. We need more ways to indirectly measure or predict these parameters from the already available data. However, the predictions are extremely sensitive to the extent of the lithology, the amount of the available data and local geology. With the application of

machine learning, we can indirectly predict the values of these geomechanical parameters using the available data of phase velocity and confining pressure. There is a need for a more robust and sophisticated model that can predict these parameters with fewer error margins. The use of these techniques may reduce costs for the direct measurement tests of the geomechanical parameters. Using machine learning can be a reliable technology as it gives fair accuracy for predicting these parameters.

ABSTRACT

The elastic velocities and confining pressures are instrumental in the accurate estimation of geomechanical properties (Young's modulus and Poisson's ratio) and anisotropic Thomsen parameters. Geomechanical properties and Thomsen parameters are critical in geomechanical applications such as hydraulic fracturing design, wellbore stability and rock failure analysis, in situ stress determination, and assessment of the response of reservoirs and surrounding rocks to changes in pore pressure and stress. Four different types of rock samples were considered for this study. The velocity data of P-wave, Sh-wave and Sv-wave, as a function of confining pressure and orientation, was used. With increasing confining pressure, it was observed that the Thomsen anisotropic parameters (ε, γ, δ) exhibit different trends. The three Thomsen anisotropic parameters (ε, γ, δ) and geomechanical properties, namely Young's moduli and Poisson's ratios, of the samples were estimated. The primary objective of the study is to test Machine Learning and conduct a comparative analysis with the conventional mathematical approach; to place emphasis on the use of Machine Learning and Artificial Intelligence in the Oil & Gas industry and to highlight its future potential to help in the digital transformation of the industry. Two different Machine Learning models, Ordinary Least Square method and Random Forest method, were used to predict the aforementioned geomechanical properties from the wave velocity and confining pressure data. The results depicted that the used methodologies were swift and reliable (93.5% accuracy) in the estimation of geomechanical properties and can be used in geomechanical modelling of petroleum reservoirs on the industry scale. Through this study, it has been observed that the Young's modulus and Poisson's ratio

are strongly affected by the anisotropy parameters, with this relationship being depicted through the correlation matrix generated for each rock sample.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

1. INTRODUCTION

The anisotropy in a rock is defined as its properties that vary with the direction of observation. This anisotropy is present everywhere within the subsurface. It is the different geological origins of each rock which create different sedimentary features in them. This, along with a rock's stress anisotropy, results in anisotropic behavioural trends. These properties are often measured parallelly or perpendicularly to the sedimentary features (bedding planes), with the earth's anisotropic response to changes investigated with the help of seismic, sonic or ultrasonic surveys. (Barton & Quadros, 2015)

Fundamental geomechanical properties include stress, strain, Poisson's ratio, Young's modulus, compressive strength. Geomechanical evaluation is required in petroleum engineering for rock failure prediction, analysis of wellbore stability, in-situ stress determination, hydraulic fracturing design, and anisotropy measurement. Geomechanical rock properties are a subset of petrophysical parameters that can be calculated directly in rock mechanics labs or field experiments (Schön, 2011). However, since they are more or less highly correlated with other petrophysical parameters (e.g., elastic wave velocities), an "indirect" derivation from geophysical measurements is being researched and applied (Schön, 2011).

Thomsen (1986) proposed three parameters to characterise anisotropy, in addition to the normal V_P, V_S, and ρ. Thomsen's parameters include ϵ (epsilon), γ (gamma), and δ (delta). The short offset effect, δ or delta , captures the relationship between the velocity required to flatten gathers (the NMO velocity) and the zero-offset average velocity calculated by check shots. It is simple to calculate, but it may be challenging to comprehend physically. According to

Thomsen, the long offset effect, also known as ϵ or epsilon, is the fractional difference between vertical and horizontal P velocities, or the parameter usually referred to as a rock's anisotropy. Horizontal velocity, on the other hand, is difficult to measure. The shear wave effect, also known as gamma, contrasts a horizontal shear wave with horizontal polarisation to a vertical shear wave. He said, "Using sonic technology, it is possible to decide in a single well. As a result, the relationship between and is of great importance." (Hall, 2015)

The effects of anisotropy in seismic data will reveal a lot about the Earth's processes and mineralogy. Seismic anisotropy has gotten a lot of attention from academia and industry in the last two decades, thanks to advances in anisotropy parameter estimation, the transition from post-stack imaging to pre-stack depth migration, and the wider offset and azimuthal coverage of 3D surveys. Anisotropic models are currently used in many seismic processing and inversion processes, offering a significant improvement in resolution and efficiency of seismic imaging. The use of an anisotropy velocity model in combination with seismic imaging has greatly reduced the uncertainty surrounding internal and bounding-fault sites, reducing the likelihood of making an investment decision solely based on seismic interpretation.

In addition, the discovery of a connection between anisotropy parameters, fracture orientation, and density has resulted in the creation of practical reservoir characterization techniques. If fractures are considered during the drilling decision process, the drainage area of each production well can be substantially increased, thanks to the acquisition of such information as fracture spatial distribution and density. The drilling cost of exploration and production (E&P)

projects will be greatly reduced because there will be less wells due to the increased drainage area.

One of the most critical aspects of preparing a strategy for hydraulic fracking is understanding the geomechanical rock properties. It's critical to reduce operational risk and maximise production while spending as little money as possible, especially in ultra-tight complex formations like shale, where operational risk is high due to formation uncertainty. As a result, machine learning and artificial intelligence (AI) are increasingly important in the oil industry, as they produce precise information by integrating log and core data. This method is important for predicting the geomechanical properties of shale, the most heterogeneous rock with the least desirable wettability for hydrocarbon flow. As a result, machine learning and artificial intelligence (AI) can combine and compare data more accurately than a human can. Manual integration and correlation are less accurate, efficient, and, most importantly, costly and time-consuming. Many researchers have concentrated on combining core and log data to calculate geomechanical properties and performance sweet spots in the Eagle Ford and Barnett formations using machine learning and artificial intelligence. (Syed, AlShamsi, Dah, & Neghabhan, 2020)

In the geomechanical analysis of petroleum reservoirs, wave velocities, Poisson's ratio, Young's, shear, and bulk modulus are all important rock mechanical properties. Direct measurement of these parameters is usually difficult, particularly in older wells, due to the high cost of testing or a lack of available data. Hence, to predict these parameters from available data, indirect methods are frequently used. The most basic and commonly used method is

empirical equations. These relationships, on the other hand, are extremely susceptible to different forms of fluids or lithologies, and are often unrelated to local geology. In recent years, intelligent systems have been used in a variety of fields of science and technology, and they have consistently proven to be useful in prediction and optimization problems. (Rajabi & Tingay, 2013)

Challenges that the industry currently faces include complex mathematical modelling that is extremely difficult and time-consuming. Apart from that, several complex calculations are performed to calculate the stiffness constant using the Voigt matrix, which can have dimensions up to 81 x 81. The time needed to solve the matrix of such large dimensions is extraordinarily complex and takes several months for the solution to be reached. These challenges can be easily overcome using AI/ML to calculate geomechanical and anisotropic parameters, giving reasonably high accuracy.

Data derived from lab study of four types of cores, Saturated Sandstone, Dry Sandstone, Sandy Shale and Shale will be used in the current project. The aim of this project is to determine how elastic anisotropy affects Young's modulus and Poisson's ratio. V_P, V_{Sh} and V_{SV} wave velocity data measured as a function of confining pressure, obtained through the ultrasonic transmission method was used for determination of geomechanical properties and anisotropy parameters. The aim of the project is to use a data-driven approach and machine learning algorithms to predict certain geomechanical parameters.

2. METHODOLOGY

2.1 Mathematical Analysis

Core data (velocity and density) of four different type of sedimentary rock is taken to calculate the Thomsen's anisotropic parameters and geomechanical property with respect to different confining pressure. Cores of different rock types were used for the study are dry sandstone, saturated sandstone, sandy shale and shale from a particular basin.

Generally, sedimentary rocks behaves as a transverse isotropic material. For transverse isotropic material, each layer has the same properties in horizontal direction but different properties in the vertical direction. The horizontal layer is the plane of isotropy and the vertical axis is the axis of symmetry. In vertical transverse isotropy, the horizontal properties is different from the vertical properties.

For measurement of transverse isotropy, standard three plug method is used. By using standard three plug method, each rock sample is cut in three different orientations which are parallel, perpendicular and typically 45˚ to the axis of symmetry (Wang, 2002). Phase velocity measurement for each core sample by one compressional (P-wave) and two orthogonal shear waves (S-wave) is performed using ultrasonic laboratory measurements. So, for each core sample total three velocity are determined therefore for each rock sample total nine velocity is determined.

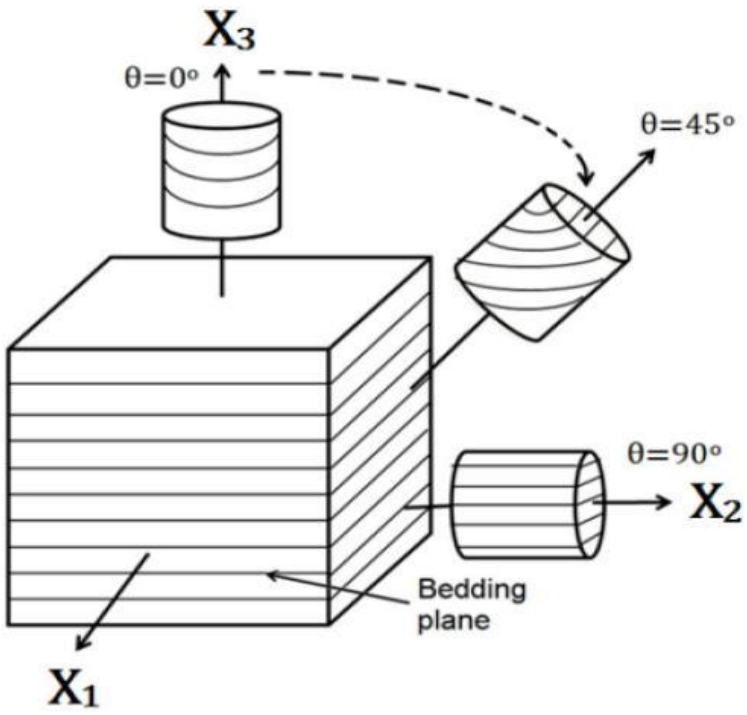

Figure 2.1: Estimation of Geomechanical and Anisotropic Parameters using Core Data: An Artificial Intelligence and Machine Learning Approach (Wang, 2002)

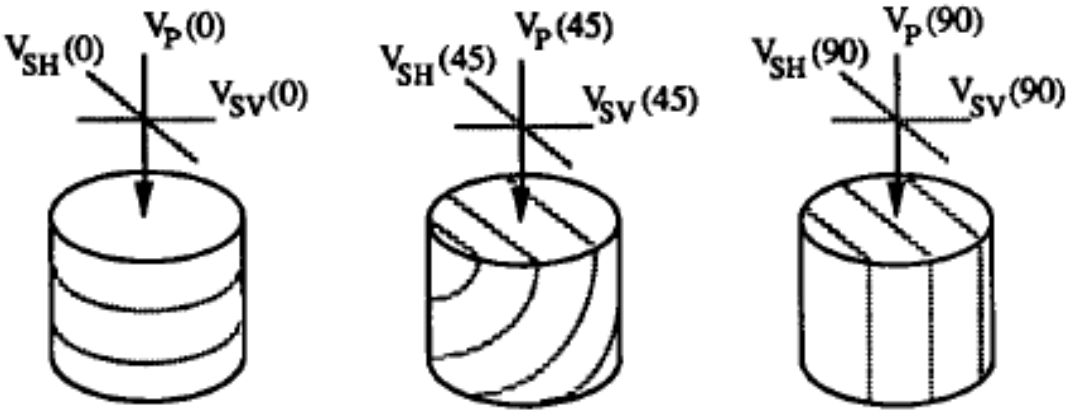

Figure 2.2: Estimation of Geomechanical and Anisotropic Parameters using Core Data: An Artificial Intelligence and Machine Learning Approach (Vernik & Nur, 1992)

The following equation expressed the relationship between stress(σ_{ij}) and strain(C_{ijkl}) in vertical transverse isotropy (VTI):

$$\sigma_{ij} = C_{ijkl} e_{kl} \tag{2-1}$$

where C_{ijkl} is the Voigt matrix (Mavko, Mukerji, & Dvorkin, 1998). If vertical axis is denoted by Z, then other two principal axes (X and Y) are parallel to the transversely isotropic plane.

In linear elasticity, the stress and strain relationship is described by Hooke's law, i.e.,

$$\sigma = Ce \tag{2-2}$$

Or using Voigt notation,

$$\begin{bmatrix}\sigma_1\\ \sigma_2\\ \sigma_3\\ \sigma_4\\ \sigma_5\\ \sigma_6\end{bmatrix} = \begin{bmatrix}C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16}\\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26}\\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36}\\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & C_{46}\\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} & C_{56}\\ C_{61} & C_{62} & C_{63} & C_{64} & C_{65} & C_{66}\end{bmatrix}\begin{bmatrix}e_1\\ e_2\\ e_3\\ e_4\\ e_5\\ e_6\end{bmatrix} \tag{2-3}$$

For Vertical isotropic material, this Voigt matrix is reduced to six elastic constants (five independent and one is dependent constants, which are C_{11}, C_{12}, C_{33}, C_{44}, C_{66} and C_{13}.

$$C = \begin{bmatrix} C_{11} & C_{11}-2C_{66} & C_{13} & 0 & 0 & 0\\ C_{11}-2C_{66} & C_{11} & C_{13} & 0 & 0 & 0\\ C_{13} & C_{13} & C_{33} & 0 & 0 & 0\\ 0 & 0 & 0 & C_{44} & 0 & 0\\ 0 & 0 & 0 & 0 & C_{55} & 0\\ 0 & 0 & 0 & 0 & 0 & C_{66}\end{bmatrix} \tag{2-4}$$

From equation, Voigt matrix C, there are five non-zero independent elastic constants: C_{11}, C_{44}, C_{66}, C_{33}, and C_{13}. The sixth elastic constant is $C_{12} = C_{11} - 2C_{66}$, where,

C_{11} = in-plane compressional modulus,

C_{33} = out-of-plane compressional modulus,

C_{44} = out-of-plane shear modulus,

C_{66} = the in-plane shear modulus,

C_{13} = important constant that controls the shape of the wave surfaces.

For calculation of geomechanical parameters and Thomsen anisotropic parameters, we need to find the value of five non-zero independent elastic constants. For calculation of five non-zero independent elastic constants, we need to use the compliance matrix which is an inverse of the elastic stiffness matrix. By solving the compliance matrix, elastic constants are calculated.

One vertical Young's modulus (E3) and one horizontal Young's modulus (E1) and two dynamic Poisson's ratios can be expressed as a function of five elastic stiffness Cij for a hexagonal material (VTI material) as follows (Pena, 1998):

$$E_1 = \frac{[C_{33}(C_{11} + C_{12}) - 2C_{13}^2](C_{11} - C_{12})}{C_{33}C_{11} - C_{13}} \quad 2\text{-}5$$

$$E_3 = \frac{[C_{33}(C_{11} + C_{12}) - 2C_{13}^2]}{C_{11} + C_{12}} \quad 2\text{-}6$$

$$\vartheta_{31} = \frac{C_{13}}{C_{11} + C_{12}} \quad 2\text{-}7$$

$$\vartheta_{12} = \frac{C_{33}C_{12} - 2C_{13}^2}{C_{11}C_{33} - 2C_{13}^2} \quad 2\text{-}8$$

These dynamic Poison's ratios ϑ_{ij}, are the ratio of the transverse to longitudinal strains when the uniaxial stress is applied in the same direction of longitudinal strain.

Anisotropic parameters can be expressed by the following mathematical equations suggested by Thomsen (1986) as a function of five elastic stiffness C_{ij}:

$$\epsilon = \frac{C_{11} - C_{33}}{2C_{33}} \quad (2\text{-}9)$$

$$\delta = \frac{(C_{13} + C_{44})^2 - (C_{33} - C_{44})^2}{2C_{33}(C_{33} - C_{44})} \quad (2\text{-}10)$$

$$\gamma = \frac{C_{66} - C_{44}}{2C_{44}} \quad (2\text{-}11)$$

2.2 Machine Learning Application

2.2.1 Data Preparation

For machine learning application, once the measurements data from the core samples are collected there is a complete procedure which must be carried out before training the machine learning model (Brownlee, How to Prepare Data For Machine Learning, 2013):

1. **Data Selection**

The aim of this step is to pick a subset of all available data. In this step, *pandas* and *numpy* library of python are used. In this process, a most considerable data has been selected which is best suitable for our problem statement. (Brownlee, How to Prepare Data For Machine Learning, 2013)

2. **Data Pre-Processing**

Three common data pre-processing steps include:

- *Formatting:* The data that has been chosen may not be in a workable format. As a result, the data has to be converted to an excel spreadsheet. (Brownlee, How to Prepare Data For Machine Learning, 2013)

- *Cleaning:* The removal or correction of missing data is known as data cleaning. It's possible that certain data instances are missing, and they'll need to be deleted. Additionally, some of the attributes may contain confidential information, and these attributes may need to be anonymized or deleted completely from the data. The *isnull()* is a function for the null value of the data set, and the outlier removal method has been used to remove outliers from the data. (Brownlee, How to Prepare Data For Machine Learning, 2013)
- *Sampling:* It's possible that there's a lot more selected data than is needed. More data can lead to much longer algorithm execution times as well as increased processing and memory requirements. As a result, before considering the entire dataset, a smaller representative subset of the selected data works much better for exploring and prototyping solutions. (Brownlee, How to Prepare Data For Machine Learning, 2013)

3. Data Transformation

Scaling, attribute decompositions, and attribute aggregations are three common data transformations. Feature engineering is another term for this move. (Brownlee, How to Prepare Data For Machine Learning, 2013)

- *Scaling:* The pre-processed data could include attributes with a mix of scales for different quantities including dollars, kilogrammes, and sales volume. Many machine learning algorithms prefer data attributes with the same scale, such as 0 to 1 for the

smallest and largest value for a given element. Consider any feature scaling that may be needed. (Brownlee, How to Prepare Data For Machine Learning, 2013)

- *Decomposition:* When a complex idea is broken down into its constituent parts, there may be features that are more useful to a machine learning system. A date, for example, may have day and time elements, each of which could be split further. Perhaps only the time of day matters in the solution of the problem. (Brownlee, How to Prepare Data For Machine Learning, 2013)
- *Aggregation:* There may be features that can be combined into a single function to make the issue more meaningful. For example, each time a customer logged into a system, there may be data instances for the number of logins, allowing the extra instances to be discarded. Consider the different types of function aggregations that could be used. (Brownlee, How to Prepare Data For Machine Learning, 2013)

4. Outlier Detection and Removal

Outliers are extreme values that differ significantly from the rest of the data. Outliers in a normal distribution, for example, may be values near the tails of the distribution. Outlier mining, outlier modelling, novelty detection, and anomaly detection are all terms used in data mining and machine learning to describe the process of detecting outliers. (Brownlee, How to Identify Outliers in your Data, 2013)

A few outlier detection methods are given below:

- *Extreme Value Analysis:* Calculate the statistical tails of the data's underlying distribution. For univariate results, for example, statistical methods such as z-scores are used. (Brownlee, How to Prepare Data For Machine Learning, 2013)
- *Probabilistic and Statistical Models:* Using a probabilistic model of the results, find the most unlikely cases. Gaussian mixture models, for example, benefit from expectation-maximization optimization. (Brownlee, How to Identify Outliers in your Data, 2013)
- *Linear Models:* Linear correlations are used in projection methods to model data into lower dimensions. Outliers include significant residual errors in data and principal component analysis (PCA). (Brownlee, How to Identify Outliers in your Data, 2013)
- *Proximity-based Models:* Cluster, density, or nearest neighbour analysis are used to separate data instances from the rest of the data. (Brownlee, How to Identify Outliers in your Data, 2013)
- *Information Theoretic Models:* Outliers are data instances that add to the complexity of a dataset (minimum code length). (Brownlee, How to Identify Outliers in your Data, 2013)
- *High-Dimensional Outlier Detection:* Distance-based measurements in higher dimensions are broken down using methods that scan subspaces for outliers (curse of dimensionality). (Brownlee, How to Identify Outliers in your Data, 2013)

2.2.2 Data Splitting

A dataset is divided into two subsets as part of the process. The training dataset is the first subset of data that is used to match the model. The second subset is not used to train the model; instead, the model is fed the dataset's input variable, which is then used to make predictions and compare them to the expected values. The test dataset is the name given to the second collection of results. (Brownlee, Train-Test Split for Evaluating Machine Learning Algorithms, 2020)

- *Train Dataset:* Used to fit the machine learning model
- *Test Dataset:* Used to evaluate the fitted machine learning model and to check for accuracy

The size of the train and test sets, which is usually expressed as a percentage between 0 and 1 for either the train or test datasets, is the procedure's key configuration parameter. For example, a training set with a size of 0.67 (67%) is allocated to the test set with the remaining percentage of 0.33 (33%). (Brownlee, Train-Test Split for Evaluating Machine Learning Algorithms, 2020) Here, using the *train_test_split()* syntax of python, the data has been split into the train data set and test dataset.

2.2.3 Dummy Data

The amount of available original data is not sufficient to train the machine learning model, so instead, we created dummy data with the help of python functions using the original data. The python function has some basic input arguments which use the minimum, maximum, mode

value, and the quantity of the data points to be generated. Suppose we want 50 dummy data points of the available original data; we need to input the fundamental above listed argument values. The function generates values that can be used better to understand the complexities of our machine learning model.

2.2.4 Data Visualization

In applied analytics and machine learning, data visualisation is a must-have skill. It can be useful for identifying trends, corrupt data, outliers, and other things while exploring and getting to know a dataset. Data visualizations can articulate and explain key relationships in plots and charts that are more visceral to yourself and stakeholders than indicators of affiliation or meaning with only a little domain information. The *matplotlib()* and *seaborn()* library of python are used for the data visualisation.

2.2.5 Machine Learning Model Selection

Once the excel file with all the features (confining pressure and shear waves in different directions) variables and output (Thomsen's parameters) variables have been imported, then comes the data pre-processing, which is a very crucial step that helps us understand the data and any flaws that are associated with it which may affect the overall accuracy of the machine learning model.

To make predictions, visualisation is an important step that gives us a graphic representation of our data. When we have very numerous data, by just looking at the numbers, we cannot interpret

anything out of it unless we visualise the data in a graphical representation. For instance, if we have some density-neutron log data, just looking at numbers would not help the geologist drive important decisions. When we put these both data into a graphical visual context, we can have a better understanding of the logs and their features (for example, the crossovers). So, visualisation helps to know the trends of the data, the patterns and outliers.

Accuracy is a crucial parameter for the selection of the machine learning model. However, our focus of the study is to predict the Thompson parameter values with given shear waves values in a different direction, and it falls under the linear regression problem. In the regression problem/model, there are specific evaluation metrics. Let have a look at some critical evaluation metrics.

1. **R-squared value (R^2):** R^2 is a statistical measure that corresponds to the magnitude of the correlation between the observed outcome and the predicted value that the model will predict. So, if a model achieves an R^2 score of 1, then it can be understood that both variables are perfectly correlated to each other, which implies no variance. In another way, (total variance explained by the model)/(total variance) the value of this equation signifies the quality of the correlation between the variables. The closer the value of R^2 to 1, the better the model is considered to be fit. (Kassambara, 2018)

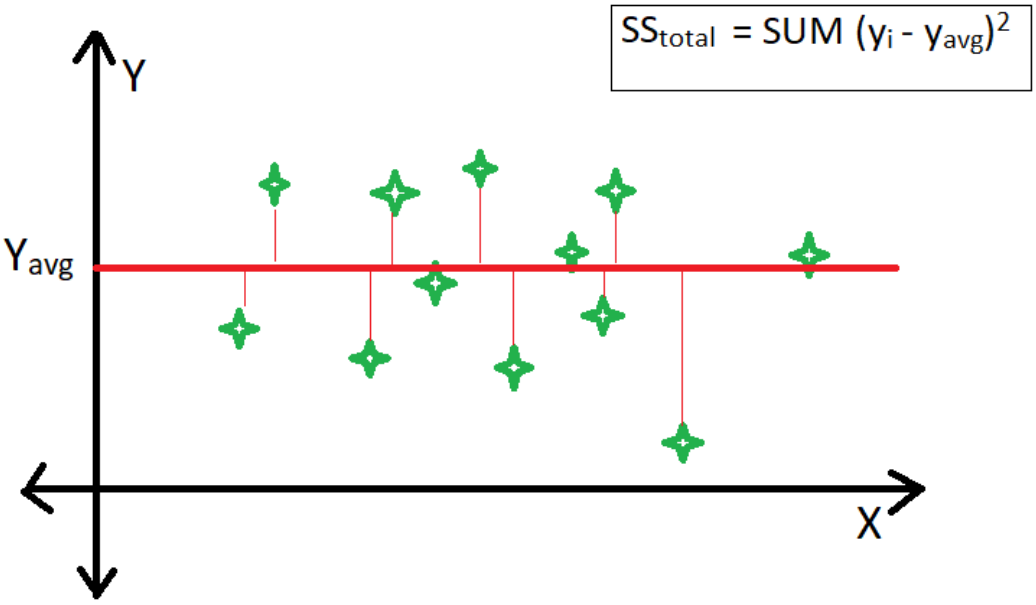

Figure 2.3: Graphical representation of R^2 value. (Jain, 2019)

2. **Adjusted R-squared:** It is an updated version of R-squared that accounts for predictors in a regression model that are no longer relevant. It shows whether adding more independent variables improves the model or not. The value of the Adjusted R-squared is always lower than the R-squared.

3. **Root Mean Squared Error (RMSE):** Another common method for calculating regression prediction errors. It's simply the average of the squared errors, or the difference between the dataset's observed value and the model's predicted value. The square root of the mean squared error (MSE), which is the average squared difference between the observed real outcome values and the values expected by the model, is the RMSE in mathematics. The better the fit, the lower the RMSE value.

4. **Mean Absolute error (MAE):** It is similar to RMSE, except that MAE tests the model's prediction error. It is the average absolute difference between observed and expected outcomes in mathematics. MAE is somewhat unaffected by outliers in the dataset.

R-squared tells us how the independent variables explain much variance in the dependent variable. In a general way, if we add more amount of the observations and more independent variables, the value of R-squared increases, but when the R-squared value does not increase any further with adding more independent variables, we need to understand that the added variables are uncorrelated with the dependent variable. In most complex predictions, it is recommended to use Adjusted R-squared in place of R-squared for model evaluation as it gives some penalty for an extra variable if the previous variable does not explain the dependent variable more correctly. However, in our case, we have a less complex situation, so it is better to use the R-squared method for model evaluation and selection in our study.

2.2.6 Algorithms used in Machine Learning

1. Ordinary Least Squares (OLS)

Ordinary least squares (OLS) is a linear least-squares approach used to estimate unknown parameters in a linear regression model. The theory of least squares is used by OLS to determine the parameters of a linear function of a set of explanatory variables: minimizing the number of squares of discrepancies between the observed dependent variable (values of the variable being observed) in the given dataset and those expected by the independent variable's linear function. (Sluijmers, 2020)

The sum of the squared distances between each data point in the set and the corresponding point on the regression surface, measured parallel to the axis of the dependent variable—the smaller the differences, the better the model fit the data. In a simple linear regression, where there is

only one regressor on the right side of the regression equation, the resulting estimator can be represented using a simple formula. (Sluijmers, 2020)

The OLS coefficient estimates for the simple linear regression are as follows (Sluijmers, 2020):

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_0 \bar{x} \tag{2-12}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2-13}$$

where the "hats" above the coefficients denote coefficient estimates, and the "lines" above the x and y variables denote sample averages, which are computed as follows (Sluijmers, 2020):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2-14}$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{2-15}$$

2. Random Forest

The bagging method is used to train Random Forests. Bagging, also known as Bootstrap Aggregating, entails sampling subsets of the training data at random, fitting a model to these smaller data sets, and aggregating the predictions. Provided that sampling with replacement is used, this method allows several instances to be used repeatedly for training. Tree bagging entails selecting subsets of the training set, fitting each with a Decision Tree, and combining the results. (Reinstein, 2017)

By applying the bagging method to the feature space, the Random Forest method adds more randomness and variety. Instead of looking greedily for the best predictors to establish branches, it randomly samples elements of the predictor space, increasing diversity and lowering variance while maintaining or increasing bias. This process, also known as "feature bagging," leads to a more robust model. (Reinstein, 2017)

Every new data point in the Random Forests algorithm goes through the same process as the previous one, except this time it visits all of the trees in the ensemble, which were grown using random samples of both training data and features. The functions for aggregation used can vary depending on the task at hand. It uses the mode or most frequent class predicted by individual trees (also known as a majority vote) for classification problems, and the average prediction of each tree for regression tasks. (Reinstein, 2017)

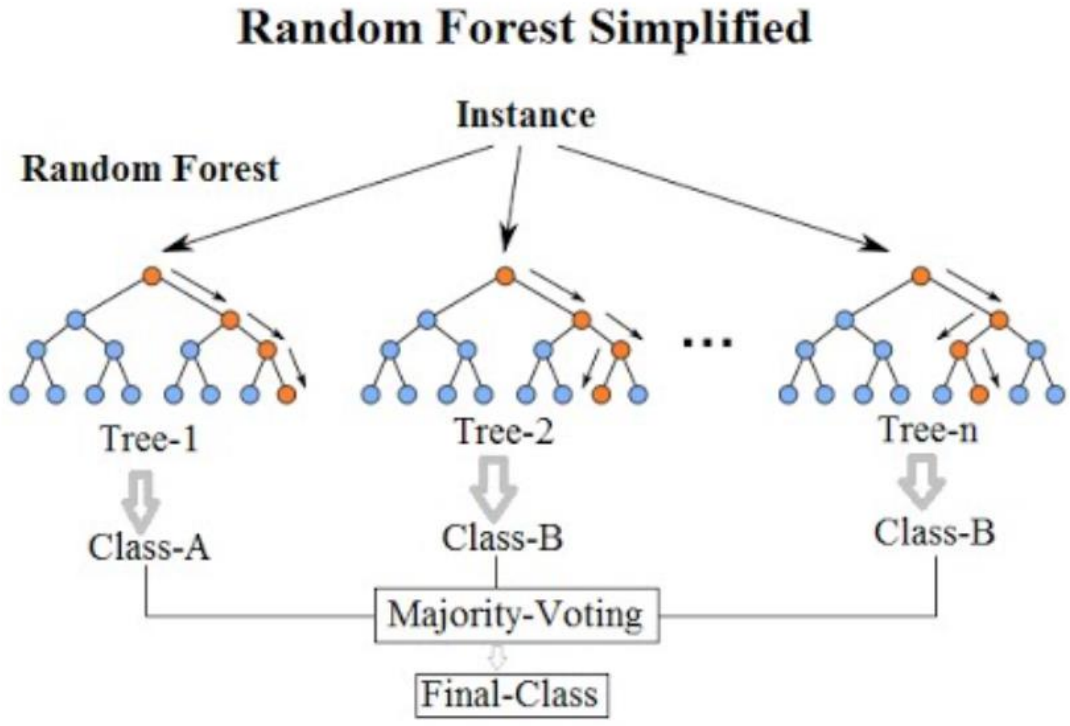

Figure 2.4: Diagrammatical representation of Random Forest method (Reinstein, 2017)

3. RESULTS AND DISCUSSION

Dry sandstone, Shale, Sandy shale and Saturated sandstone were the four types of rock samples that were considered to be appropriately representative of the anisotropic and elastic behaviour observed in a typical sedimentary column. Each of these rock types exhibit varying sensitivity to confining pressure, and consequently, the anisotropies vary as a function of it.

The anisotropy parameters are plotted. Each meaning of the Thomsen anisotropy parameters is given as follows: The first parameter is γ; from the original equation, it is evident that γ only depends on the S-wave velocity component, whether the wave velocity is fast or slow. Hence, the parameter γ describes the anisotropy condition of the S-wave velocity. When the anisotropy of S-wave increases, the value of anisotropy parameter γ increases. The second parameter is ε; ε describes anisotropy of P-wave velocity. When the anisotropy of P-wave increases, the value of anisotropy parameter ε increases. The third parameter is δ; δ describes both P-wave and S-wave. The increase in anisotropy of P-wave and S-wave shows an increase in the value of anisotropy parameter δ.

Mechanical properties of rocks, including Young's modulus (Horizontal and Vertical) and Poisson's ratio (Horizontal and Vertical), have an essential role to play in the geomechanical analysis of petroleum reservoirs. Due to a lack of available data and high testing costs, the direct measurement of these parameters is often not feasible. Therefore, the prediction of these parameters through the help of machine learning models, from what little data is available, has proven to be effective. The usage of empirical equations is the simplest method, and one that's

still prevalent. However, different types of lithologies or fluids often tend to adversely affect these empirical relationships, and as a result, they may not be relevant for local geology. To deal with this complexity, AI and machine learning have been demonstrated to help predict and optimise problems in a variety of sciences and technologies in recent years. Herein, a set of geomechanical parameters for different types of rocks has been predicted using Machine learning models (Ordinary Least Square method and Random Forest method). For this purpose, mechanical properties of rocks, belonging to different lithologies, were predicted from wave velocities measured in the experimental studies on the core. The results depicted that the used methodologies were swift and reliable (93.5% accuracy) in the estimation of geomechanical properties and can be used in geomechanical modelling of petroleum reservoirs on the industry scale.

3.1 The Effect of Confining Pressure on Thomsen Anisotropic Parameters

1. *Dry Sandstone*

With increasing confining pressure, the Thomsen anisotropic parameters show different trends. The P-wave anisotropy parameter ε increases linearly with increasing confining pressure and the S-wave anisotropy parameter γ, which shows a similar trend and increases in confining pressure. Furthermore, the third parameter is δ which describes that both P-wave and S-wave remains constant at zero and does not change with an increase in confining pressure. (Figure 3.1)

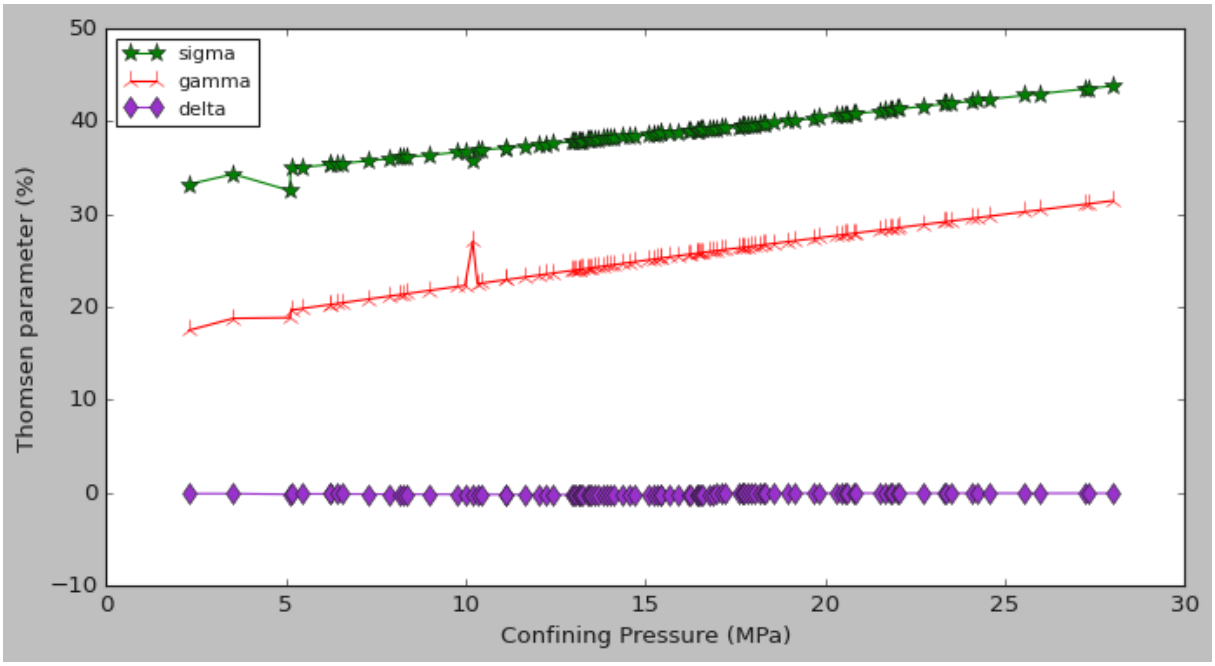

Figure 3.1: Thomsen's parameters for dry sandstone compared as a function of confining pressure.

2. *Sandy Shale*

With increasing values of confining pressure, the P-wave anisotropy parameter ε increases gradually, showing high values, while the S-wave anisotropy parameter γ also shows a similar trend, but with negative values as it approaches zero. The third anisotropy parameter δ, which describes both P-wave and S-wave, remains constant at zero and does not change. (Figure 3.2)

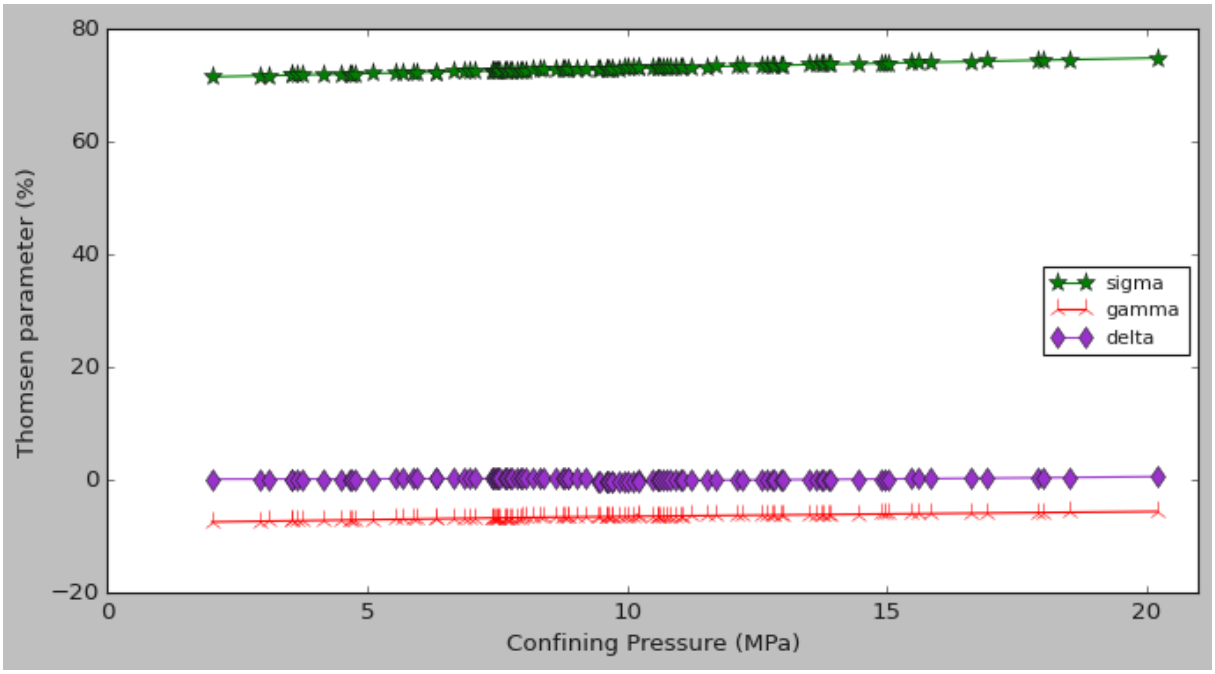

Figure 3.2: Thomsen's parameters for sandy shale compared as a function of confining pressure.

3. Shale

With increasing values of confining pressure, the P-wave anisotropy parameter ε increases gradually, showing high values, while the S-wave anisotropy parameter γ also shows a similar trend but with slightly lower values. The third anisotropy parameter δ, which describes both P-wave and S-wave, remains constant at zero and does not change. (Figure 3.3)

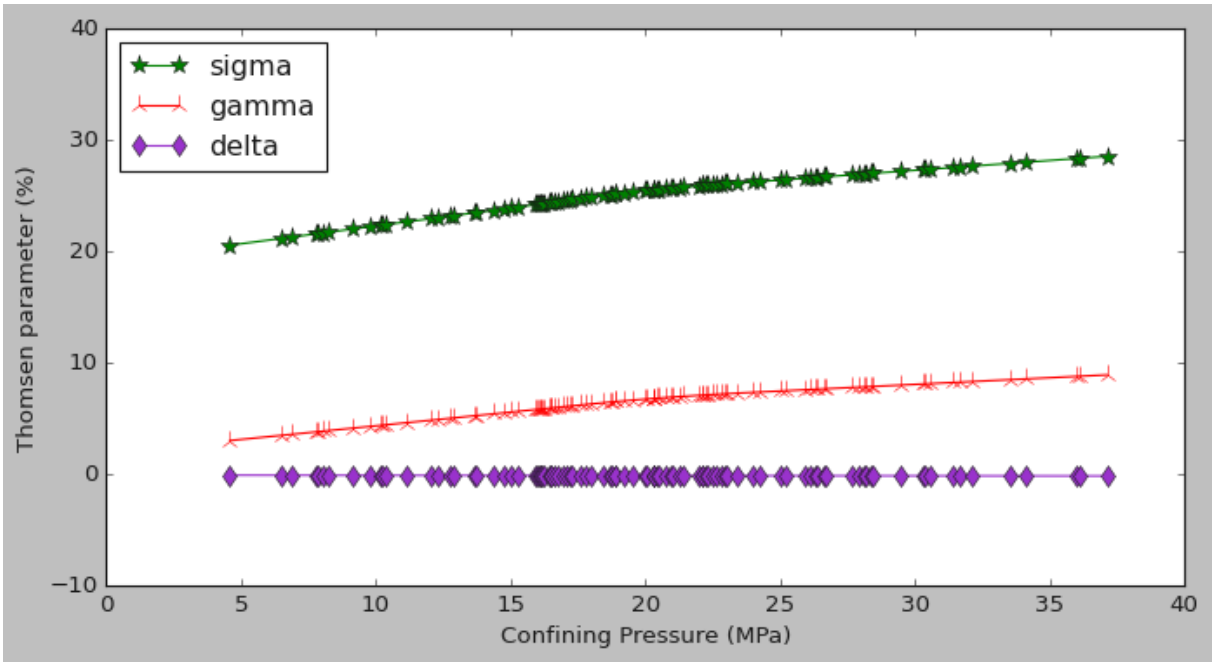

Figure 3.3: Thomsen's parameters for shale compared as a function of confining pressure.

4. Saturated Sandstone

With increasing values of confining pressure, the P-wave anisotropy parameter ε shows a relatively sharper decline before becoming more gradual. In contrast, the S-wave anisotropy parameter γ moves in the opposite direction, remaining constant initially before showing an upward trend. The third anisotropy parameter δ, which describes both P-wave and S-wave, remains constant at zero and does not change. (Figure 3.4)

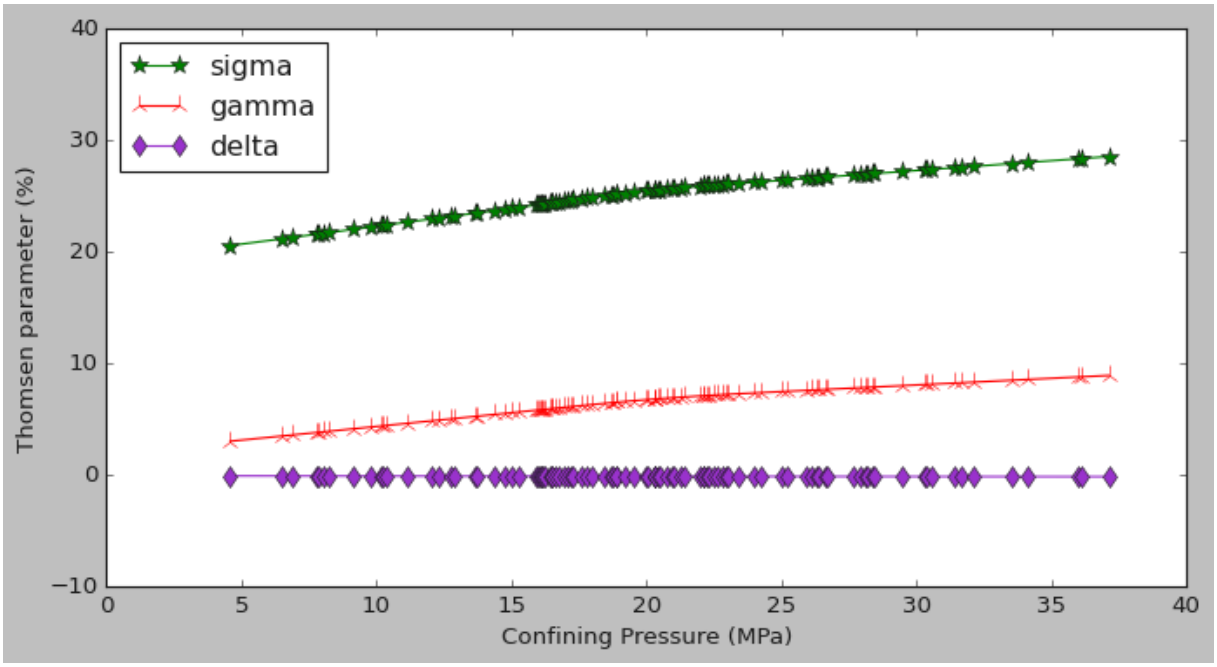

Figure 3.4: Thomsen's parameters for saturated sandstone compared as a function of confining pressure.

3.2 Estimation of Geomechanical Parameters using Predictive Modelling Techniques and Comparison with Mathematical Model

E_1 (horizontal young's modulus) is an indirect measure of the ratio of stress to axial strain when the specimen is under uniaxial stress in the X-direction. E_3 (vertical young's modulus) is an indirect measure of the ratio of stress to axial strain when the specimen is under uniaxial stress in the Z-direction, typical to the plane isotropy.

For all types of rocks except for saturated sandstone, $E_1 > E_3$. This indicates that when equal stress is applied in both directions, more strain results in the vertical direction compared to the horizontal direction because young's modulus is inversely proportional to the strain. This reflects the vertical isotropic behaviour of the material.

For saturated sandstone, there is no significant difference between the horizontal and vertical Young's Modulus. Moreover, since the difference between horizontal and vertical Young's Modulus is nearly zero for an isotropic material, it can be inferred that a saturated sandstone behaves like an isotropic material.

V_{12} (horizontal Poisson's ratio) is an indirect measure of lateral to axial strain ratio when the specimen is under uniaxial stress in the X-direction.

The value of ϑ_{12} being high and almost constant indicates that the rock is very stiff and linearly elastic in the XY plane because of the absence of voids and the discontinuities in this plane.

V_{31} (vertical Poisson's ratio) is an indirect measurement of the ratio of lateral strain to axial strain when the sample is subjected to uniaxial stress in the Z-direction, normal to the plane of isotropy.

For all types of rocks except for saturated sandstone, $V_{12} > V_{31}$. This means that the specimen is softer in the vertical direction than in the horizontal direction. This reflects the vertical isotropic behaviour of the material.

Two different Machine Learning models, Ordinary Least Square (OLS) and Random Forest (RF), were used to predict horizontal & vertical Young's Modulus (E_1 & E_3) and horizontal & vertical Poisson's Ratio (V_{12} & V_{31}) for the below samples. The predicted values were then compared with the values from the mathematical model, calculated using empirical equations, to determine the accuracy.

1. *Dry Sandstone*

While calculating E_1 and E_3 for a given set of confining pressures through the OLS method and Random Forest method, the OLS method gives an accuracy of approximately 98% and 94%, respectively, compensating the number of predictions only three out of six input values. Comparing it to the Random Forest method gives an accuracy of 92% and 98%, respectively, to predict the outcome for all six input values. (Table 3-1)

While calculating V_{12} and V_{31} for a given set of confining pressures through the OLS method and Random Forest method, the OLS method gives an accuracy of approximately 98% and 75%, respectively, compensating the number of predictions only three out of six input values. Comparing it to the Random Forest method gives an accuracy of 89% and 74%, respectively, to predict the outcome for all six input values in the first case. (Table 3-2)

For this particular sample, the RF method was vastly superior to the OLS method, and the obtained values were pretty accurate compared to those of the mathematical model.

Table 3-1: Comparison between mathematical and predicted model for estimating Young's modulus in a dry sandstone sample

Dry Sandstone										
Confining pressure (MPa)	**E_1 (Horizontal Young's Modulus (MPa))**					**E_3 (Vertical Young's Modulus (MPa))**				
	Mathematical Model	**Predicted Model**				**Mathematical Model**	**Predicted Model**			
		OLS Method	**Accuracy**	**RF Method**	**Accuracy**		**OLS Method**	**Accuracy**	**RF Method**	**Accuracy**
2.3	3.708012529	3.910	95%	3.79	98%	1.410995664	1.735	81%	1.4260	99%
5.1	3.210830011	3.216	100%	4.431	72%	1.402640235	1.4	100%	1.4240	99%
10.2	3.632629248	3.668	99%	3.955	92%	1.490805682	1.494	100%	1.5137	98%
20	5.465012748			5.67	96%	1.797465001			1.7470	97%
30	6.11027495			5.95	97%	1.838503204			1.7900	97%
30.2	6.384496232			6.12	96%	1.853176489			1.8140	98%
			Avg. Accuracy		**Avg. Accuracy**			**Avg. Accuracy**		**Avg. Accuracy**
			98%		92%			94%		98%

Table 3-2: : Comparison between the mathematical and predicted model for estimating Poisson's ratio in a dry sandstone sample

Dry Sandstone										
Confining pressure (MPa)	**V_{31} (Vertical Poisson's ratio)**					**V_{12} (Horizontal Poisson's ratio)**				
	Mathematical Model	**Predicted Model**				**Mathematical Model**	**Predicted Model**			
		OLS Method	**Accuracy**	**RF Method**	**Accuracy**		**OLS Method**	**Accuracy**	**RF Method**	**Accuracy**
2.3	-0.016860547	0.0049	29%	-0.0340	50%	0.169340116	0.1780	95%	0.162	96%
5.1	-0.135471039	-0.1390	97%	-0.1220	90%	0.100161802	0.1000	100%	0.118	85%
10.2	-0.159511085	-0.1550	97%	-0.1300	81%	0.080528974	0.0810	99%	0.103	78%
20	0					0.176851374			0.189	94%
30	0					0.233715971			0.192	82%
30.2	0					0.198308787			0.195	98%
			Avg. Accuracy		**Avg. Accuracy**			**Avg. Accuracy**		**Avg. Accuracy**
			75%		74%			98%		89%

2. Shale

While calculating E_1 and E_3 for a given set of confining pressures through the OLS method, approximately 99% was obtained for both cases. The predicted values were found to be almost identical to those calculated using the mathematical model. RF method was not considered for the given Shale sample because the OLS method was deemed accurate enough. (Table 3-3)

While calculating V_{12} and V_{31} for a set of confining pressures through the OLS method, it gives an accuracy of approximate 98% and 97%, respectively, which is almost identical to those calculated using the mathematical model. (Table 3-4)

RF method was not considered for the given Shale sample because the OLS method was deemed accurate enough.

Table 3-3: Comparison between the mathematical and predicted model for estimating Young's modulus in a shale sample

Shale						
Confining pressure (MPa)	**E_1 (Horizontal Young's Modulus (MPa))**			**E_3 (Vertical Young's Modulus (MPa))**		
	Mathematical Model	**Predicted Model**		**Mathematical Model**	**Predicted Model**	
		OLS Method	**Accuracy**		**OLS Method**	**Accuracy**
3.1	9.733428376	9.710	100%	6.003302606	6.010	100%
5.2	11.0755007	11.160	99%	6.158448828	6.316	98%
10.3	12.22255047	11.870	97%	6.341019392	6.390	99%
20.1	12.83280877	12.960	99%	6.503649957	6.411	99%
30.6	13.44198011	13.200	98%	6.491447506	6.594	98%
40.3	13.22874197	13.220	100%	6.620959109	6.624	100%
			Avg. Accuracy			**Avg. Accuracy**
			99%			99%

Table 3-4: Comparison between the mathematical and predicted model for estimating Poisson's ratio in a shale sample

Shale						
Confining pressure (MPa)	**V_{31} (Vertical Poisson's ratio)**			**V_{12} (Horizontal Poisson's ratio)**		
	Mathematical Model	**Predicted Model**		**Mathematical Model**	**Predicted Model**	
		OLS Method	**Accuracy**		**OLS Method**	**Accuracy**
3.1	0.129480576	0.129	100%	0.427600388	0.428	100%
5.2	0.104723742	0.112	94%	0.449112072	0.476	94%
10.3	0.073195914	0.068	93%	0.430831071	0.429	100%
20.1	0.051288404	0.050	97%	0.445622109	0.431	97%
30.6	0.054635724	0.054	99%	0.462438597	0.471	98%
40.3	0.058807873	0.058	99%	0.459837903	0.460	100%
			Avg. Accuracy			**Avg. Accuracy**
			97%			98%

3. Sandy Shale

While calculating E_1 and E_3 for a given set of confining pressures through the OLS method and Random Forest method, the OLS method gives an accuracy of 98% for both cases, but with the slight drawback of accurately predicting outcomes, only three of the five input values. This irregularity is due to complexity in calculating the constant elastic $C_{13,}$ which becomes a complex number at the associated pressures. On the other hand, the RF method faces no such complexities and can predict 97% and 99% accuracy for E_1 and $E_{3,}$ respectively. (Table 3-5)

While calculating V_{12} and V_{31} for a given set of confining pressures, the OLS method gives an accuracy of 68% and 78%, respectively, while the Random Forest method gives an accuracy of 99% and 74%, respectively. In V_{31} however, both the ML models were able to predict outcomes for only two of the five input values. This complication is due to the issue of data redundancy. (Table 3-6)

For this particular sample, the RF method was vastly superior to the OLS method, and the obtained values were reasonably accurate compared to those of the mathematical model.

Table 3-5: Comparison between the mathematical and predicted model for estimating Young's modulus in a sandy shale sample

Sandy Shale										
Confining pressure (MPa)	**E_1 (Horizontal Young's Modulus (MPa))**					**E_3 (Vertical Young's Modulus (MPa))**				
	Mathematical Model	**Predicted Model**				**Mathematical Model**	**Predicted Model**			
		OLS Method	**Accuracy**	**RF Method**	**Accuracy**		**OLS Method**	**Accuracy**	**RF Method**	**Accuracy**
1.3	5.566921833	5.563	100%	5.66	98%	1.32	1.31	99%	1.322	100%
3.3	5.746127921			5.405	94%	1.33			1.3	98%
5.2	5.906146999			5.583	95%	1.33			1.311	98%
10.2	5.525406037	5.536	100%	5.446	99%	1.33	1.33	100%	1.327	100%
20.2	7.043371391	7.437	95%	6.837	97%	1.36	1.43	95%	1.36	100%
			Avg. Accuracy		**Avg. Accuracy**			**Avg. Accuracy**		**Avg. Accuracy**
			98%		97%			98%		99%

Table 3-6: Comparison between the mathematical and predicted model for estimating Poisson's ratio in a sandy shale sample

Sandy Shale										
Confining pressure (MPa)	**V_{31} (Vertical Poisson's ratio)**					**V_{12} (Horizontal Poisson's ratio)**				
	Mathematical Model	**Predicted Model**				**Mathematical Model**	**Predicted Model**			
		OLS Method	**Accuracy**	**RF Method**	**Accuracy**		**OLS Method**	**Accuracy**	**RF Method**	**Accuracy**
1.3	0					0.53	0.531	100%	0.53	99%
3.3	0					0.52	-2.040	25%	0.51	98%
5.2	0					0.51	2.818	18%	0.50	99%
10.2	-0.036633466	-0.04	92%	-0.04	83%	0.49	0.490	99%	0.49	99%
20.2	0.032063781	0.05	64%	0.02	65%	0.49	0.500	97%	0.49	99%
			Avg. Accuracy		**Avg. Accuracy**			**Avg. Accuracy**		**Avg. Accuracy**
			78%		74%			68%		99%

4. Saturated Sandstone

While calculating E_1 and E_3 for a given set of confining pressures, the OLS method accuracy is approximately 68% and 66%, respectively. Comparing it to the Random Forest method gives an accuracy of 92% and 94%, respectively. (Table 3-7)

While calculating V_{12} and V_{31} for a given set of confining pressures, the OLS method gives an accuracy of approximately 59% and 57%, respectively. Comparing it to the Random Forest method gives an accuracy of 88% and 86%, respectively. (Table 3-8)

For this particular sample, the RF method was found to be vastly superior to the OLS method, and the obtained values were found to be reasonably accurate when compared to those of the mathematical model.

Table 3-7: Comparison between the mathematical and predicted model for estimating Young's modulus in a saturated sandstone sample

Saturated Sandstone										
Confining pressure (MPa)	**E_1 (Horizontal Young's Modulus (MPa))**					**E_3 (Vertical Young's Modulus (MPa))**				
	Mathematical Model	**Predicted Model**				**Mathematical Model**	**Predicted Model**			
		OLS Method	**Accuracy**	**RF Method**	**Accuracy**		**OLS Method**	**Accuracy**	**RF Method**	**Accuracy**
6.7	2.60	2.600	100%	2.562	98%	1.97	1.9660	100%	2.0270	97%
11.3	2.35	5.032	47%	2.525	93%	2.54	5.4240	47%	2.1120	83%
15.6	1.89	0.540	29%	2.502	75%	1.92	0.4087	21%	2.2071	87%
20.1	2.32	1.059	46%	2.506	92%	2.27	0.9625	42%	2.2687	100%
30.3	2.36	2.737	86%	2.523	93%	2.28	2.6780	85%	2.3032	99%
40.2	2.54	2.537	100%	2.5327	100%	2.33	2.3340	100%	2.3033	99%
			Avg. Accuracy		**Avg. Accuracy**			**Avg. Accuracy**		**Avg. Accuracy**
			68%		92%			66%		94%

Table 3-8: Comparison between the mathematical and predicted model for estimating Poisson's ratio in a saturated sandstone sample

Saturated Sandstone										
Confining pressure (MPa)	**V_{31} (Vertical Poisson's ratio)**					**V_{12} (Horizontal Poisson's ratio)**				
	Mathematical Model	**Predicted Model**				**Mathematical Model**	**Predicted Model**			
		OLS Method	**Accuracy**	**RF Method**	**Accuracy**		**OLS Method**	**Accuracy**	**RF Method**	**Accuracy**
6.7	0.20	0.1950	100%	0.2131	92%	0.27	0.2650	100%	0.2534	96%
11.3	0.16	0.6700	23%	0.2365	66%	0.32	0.7870	41%	0.2390	74%
15.6	0.35	0.0460	13%	0.2641	74%	0.18	-0.0580	32%	0.2163	83%
20.1	0.30	0.0750	25%	0.2823	94%	0.21	0.0126	6%	0.1960	93%
30.3	0.32	0.3833	82%	0.2955	94%	0.16	0.2230	73%	0.1780	91%
40.2	0.31	0.3000	98%	0.2967	97%	0.16	0.1620	100%	0.1740	93%
			Avg. Accuracy		**Avg. Accuracy**			**Avg. Accuracy**		**Avg. Accuracy**
			57%		86%			59%		88%

3.3 Relationship Between Geomechanical and Thomsen Parameters

The application of a correlation matrix interprets the relationship between the Geomechanical parameters and Thomsen parameters. An absolute value of 1 in the correlation table indicates a perfect positive linear relationship between the variables, and a correlation value close to 0 indicates no linear relationship between the variables. The sign of correlation, either positive or negative, shows the direction of the relationship. If the variables are likely to decrease or increase together, then the coefficient is positive. Similarly, if one variable increases concerning a decrease in the other variable, then there is a negative correlation and coefficient if negative.

1. *Dry Sandstone*

For a dry Sandstone sample, ε and γ are positively correlated to E_1, E_3 and V_{12} to a great degree, but they show a high negative correlation with V_{31}. On the contrary, δ shows the opposite trend whereby it exhibits a high negative correlation with E_1, E_3 and V_{12} whereas V_{31} is positively correlated to δ. (Table 3-9)

Table 3-9: Correlation between geomechanical and anisotropic Thomsen parameters in a dry sandstone sample

Dry Sandstone				
	E_1	**E_3**	**V_{12}**	**V_{31}**
ε	0.9997	0.9987	0.9994	-0.9999
γ	0.9987	0.9972	0.9985	-0.9998
δ	-0.9998	-0.9997	-0.9988	0.9989

2. Shale

For a Shale sample, ε is positively correlated to E_1, E_3 and V_{12} to a great degree, but it shows a high negative correlation with V_{31}. In the case of γ, it shows a high positive correlation with E_3 and V_{12}, but it exhibits a high negative correlation with E_1 and V_{31}. On the contrary, δ shows the opposite trend whereby it exhibits a high negative correlation with E_3 and V_{12} whereas E_1 and V_{31} are positively correlated to δ. (Table 3-10)

Table 3-10: Correlation between geomechanical and anisotropic Thomsen parameters in a shale sample

Shale				
	E_1	**E_3**	**V_{12}**	**V_{31}**
ε	0.9997	0.9987	0.9994	-0.9999
γ	-0.9997	0.9972	0.9985	-0.9998
δ	0.9996	-0.9997	-0.9988	0.9989

3. Sandy Shale

For a Sandy shale sample, ε and γ are positively correlated with V_{31} to a great degree, but it shows a moderately negative correlation with V_{12}. Moreover, it does not hold any relation with E_1 and E_3. On the contrary, δ shows a high correlation with all the geomechanical parameters. (Table 3-11)

Table 3-11: Correlation between geomechanical and anisotropic Thomsen parameters in a sandy shale sample

Sandy Shale				
	E_1	**E_3**	**V_{12}**	**V_{31}**
ε	-0.0844	0.0111	-0.5469	0.9844
γ	-0.1241	-0.0308	-0.5821	0.985
δ	0.985	0.8983	0.8018	0.9704

4. Saturated Sandstone

For a Saturated sandstone sample, ε is positively correlated with E_1 and V_{12} to a great degree, but it shows a high negative correlation with E_3 and V_{31}. Whereas in the case of γ, there is a stark difference as it shows a high positive correlation with E_3 and V_{31}, but it exhibits a high negative correlation with V_{12}. Moreover, γ has no relation with E_1. Similarly, δ exhibits a moderate and high negative correlation with E_1 and V_{12}, respectively, whereas E_3 and V_{31} are positively correlated to δ. (Table 3-12)

Table 3-12: Correlation between geomechanical and anisotropic Thomsen parameters in a saturated sandstone sample

Saturated Sandstone				
	E_1	**E_3**	**V_{12}**	**V_{31}**
ε	0.8515	-0.9906	0.9307	-0.985
γ	-0.1388	0.7323	-0.8713	0.7551
δ	-0.5919	0.9665	-0.997	0.9764

4. CONCLUSION

Currently, the conventional approach of selectively adopting many technologies and applying digitalisation may not be the best way forward. Instead, the industry would gain more if it pursued a transformative agenda with digitalisation as the foundation itself. A digital transformation at this stage can revolutionise not only the industry but also benefit society. A centred digital strategy and a culture of creativity and technology adoption would be required for such a transition. To fully realise the potential of digitalisation, all the enablers needed for a successful transition must be in place.

Based on the study, we can conclude that the elastic anisotropy parameters are the main factors that influence the estimation of hydrocarbon reservoir characterisation parameters. Vertical P-wave and S-wave velocities and the three anisotropy parameters are needed to estimate more accurate hydrocarbon reservoir characterisation parameters. The anisotropy parameters ε, γ and δ can be reasonably estimated through surface seismic data of good quality and high resolution. To estimate the remaining parameters, we must rely on downhole data, wireline measurements for sonic profiling, and other seismic profiling methods. The lab experiments on core samples would only help obtain empirical relationships between some of the parameters to develop the initial model.

After applying the ML algorithms, the anisotropy parameters and the geomechanical properties could be estimated with reasonable accuracy. Using the mathematical model would have required us to find out the stiffness constants first, a complication which has been eliminated

through the use of ML algorithms. This facilitates the direct estimation of geomechanical properties through velocity profile inputs.

On conducting a comparative analysis of the results obtained through the machine learning approach in this study with those obtained through the mathematical model, given by Brahma and Sircar, 2014, it was found that the ML methods came in with a more than respectable accuracy of 93.5%. The findings showed that the values of horizontal Young's Modulus (E1) were greater than that of vertical Young's Modulus (E3) for all rocks except for saturated sandstone, indicating their tendency to experience greater strain in the vertical direction. For all rocks except for saturated sandstone, the value of horizontal Poisson's Ratio (V12) was also found to be greater than that of vertical Poisson's Ratio (V31), indicating the softness of the rock in the vertical direction. Both these findings explain the vertical isotropic behaviour of these rocks. Therefore, it can be concluded that anisotropic parameters are highly correlated to the geomechanical properties due to the latter's directional dependency.

Moreover, we can also conclude that for a machine learning model to predict correct values with less margin of error, the model needs to be trained with more modelling data. The number of data points used in the training of a model has a substantial effect on its overall accuracy. So, to be able to learn and understand the complexities, patterns and relationships between given input and output variables, it requires more modelling data.

We can understand the effect of fewer modelling data points on the overall accuracy in the OLS method, where the OLS (Ordinary Least Square) model fails in predicting some data points of

the original data. It means the model does not entirely understand the relationships between the variables in fewer data.

5. Future Scope of Work

In this project, we have used the Ordinary Least Square (OLS) method and Random Forest method, which are well-known for their best accuracy in the petroleum industry and are frequently used to cope with the encountered real-time problems a better decision-making approach. This approach has also proven reasonably successful in our case, having yielded results with an overall average accuracy of 93.5%. Nevertheless, despite that, many other algorithms might prove to give better accuracy for this problem. This can be achieved through the process of optimisation.

An optimisation is the problem of finding a set of inputs to an objective function that results in a maximum or minimum function evaluation. It is a challenging problem that underlies many machine learning algorithms, from fitting logistic regression models to training artificial neural networks. There are perhaps hundreds of popular optimisation algorithms, and one particular algorithm which could prove to be successful for our problem is the Gradient Descent algorithm.

Gradient descent is a first-order iterative optimisation algorithm for finding a local minimum of a differentiable function. Gradient descent is best used when the parameters cannot be calculated analytically (e.g., using linear algebra) and must be searched for by an optimisation algorithm. This method would help reduce the margin of error in the results obtained through the OLS method and the Random Forest method. However, it is essential to note that if the execution is not appropriately done while using gradient descent, it may lead to vanishing

gradient or exploding gradient problems. These problems occur when the gradient is too small or too large, causing the algorithms to never converge.

6. REFERENCES

Barton, N., & Quadros, E. (2015). Anisotropy is Everywhere, to See, to Measure, and to Model. *Rock Mech Rock Eng, 48*, 1323–1339. doi:https://doi.org/10.1007/s00603-014-0632-7

Brahma, J., & Sircar, A. (2014). Estimation of the Effect of Anisotropy on Young's Moduli and Poisson's Ratios of Sedimentary Rocks Using Core Samples in Western and Central Part of Tripura, India. *International Journal of Geosciences, 5*, 184-195. doi:http://dx.doi.org/10.4236/ijg.2014.52020

Brownlee, J. (2013, 12 31). *How to Identify Outliers in your Data*. Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/how-to-identify-outliers-in-your-data/

Brownlee, J. (2013, 12 25). *How to Prepare Data For Machine Learning*. Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/how-to-prepare-data-for-machine-learning/

Brownlee, J. (2020, 7 24). *Train-Test Split for Evaluating Machine Learning Algorithms*. Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/

Hall, M. (2015, 2 9). *What is anisotropy?* Retrieved from Agile Scientific: https://agilescientific.com/blog/2015/2/9/what-is-anisotropy

Jain, D. (2019, 5 7). *ML | R-Squared in Regression Analysis*. Retrieved from GeeksforGeeks: https://www.geeksforgeeks.org/ml-r-squared-in-regression-analysis/

Kassambara, A. (2018, 03 11). *Regression Model Accuracy Metrics: R-square, AIC, BIC, Cp and more*. Retrieved from sthda.com: http://www.sthda.com/english/articles/38-regression-model-validation/158-regression-model-accuracy-metrics-r-square-aic-bic-cp-and-more/

Mavko, G., Mukerji, T., & Dvorkin, J. (1998). *Rock Physics Handbook: Tools for Seismic Analysis in Porous Media.* Cambridge University Press.

Pena, F. R. (1998). Elastic Properties of Sedimentary Anisotropic. *Dissertation*, 19-26. Massachusetts Institute of Technology.

Rajabi, M., & Tingay, M. (2013). Applications of Intelligent Systems in Petroleum Geomechanics - Prediction of Geomechanical Properties in Different Types of Sedimentary Rocks. *International Workshop on Geomechanics and Energy: The Ground as Energy Source and Storage*. doi:http://dx.doi.org/10.3997/2214-4609.20131949

Reinstein, I. (2017, 10). *Random Forests, Explained*. Retrieved from KDnuggets : https://www.kdnuggets.com/2017/10/random-forests-explained.html

Schön, J. H. (2011). Chapter 7 - Geomechanical Properties. In J. H. Schön, *Handbook of Petroleum Exploration and Production* (pp. 245-271). Elsevier. doi:https://doi.org/10.1016/S1567-8032(11)08007-4

Sluijmers, M. (2020, 5 25). *(Simple) Linear Regression and OLS: Introduction to the Theory*. Retrieved from Towards Data Science: https://towardsdatascience.com/simple-linear-regression-and-ols-introduction-to-the-theory-1b48f7c69867

Syed, F. I., AlShamsi, A., Dah, A. K., & Neghabhan, S. (2020). Machine Learning techniques to Model Geomechanics and Petrophysical Properties of Shale Reservoirs – A Systematic Literature Review. *Petroleum*. doi:https://doi.org/10.1016/j.petlm.2020.12.001

Thomsen, L. (1986). Weak Elastic Anisotropy. *Geophysics, 51*(10), 1954-1966. doi:http://dx.doi.org/10.1190/1.1442051

Vernik, L., & Nur, A. (1992). Ultrasonic Velocity and Anisotropy of Hydrocarbon Source Rocks. *Geophysics, 57*(5), 727-735. doi:10.1190/1.1443462

Wang, Z. (2002). Seismic Anisotropy in Sedimentary Rocks Part 1: A Single-Plug Laboratory Method. *Geophysics*, *67*(5), 1415-1422.