PREDICTION OF SONIC LOG USING PETROPHYSICAL LOGS VIA MACHINE LEARNING TECHNIQUE

AHMAD NASUHA BIN MOHAMAD SHABARI

UNIVERSITI TEKNOLOGI MALAYSIA

PREDICTION OF SONIC LOG USING PETROPHYSICAL LOGS VIA MACHINE LEARNING TECHNIQUE

AHMAD NASUHA BIN MOHAMAD SHABARI

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Petroleum Engineering

School of Chemical and Energy Engineering
Faculty of Engineering
Universiti Teknologi Malaysia

JULY 2022

ACKNOWLEDGEMENT

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor,Dr. Abdul Rahim Bin Risal, for encouragement, guidance, critics and friendship. I am also very thankful to my co-supervisor Mr Jamari Bin Mohd Shah for his guidance, advices and motivation. Without their continued support and interest, this thesis would not have been the same as presented here.

I would like to thank all my management at PETRONAS Carigali Sdn. Bhd. for giving me the opportunity to pursue this Master and for their sponsorship and also like to extend my gratitude to the team at GTS especially Tay Ai Yin for helpful discussions and support. Special thanks to MPM for granted me permission to use reservoir data as the main reference for the case study in this research.

My fellow postgraduate student should also be recognised for their support. My sincere appreciation also extends to all my colleagues and others who have provided assistance at various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all my family member.

ABSTRACT

The sonic log is the pivotal parameter for the reservoir description and fluid identification and is extensively applied in determining mechanical rock properties for rock physics, quantitative seismic interpretation, and geomechanics application. There is frequently a paucity of shear wave velocity (Vs) data in oil and gas exploration wells which is relatively due to poor borehole conditions (washout), damaged tools, offset well data, and quite expensive. This paper aims to provide a solution to predict the compressional wave (Vp) and shear wave velocities (Vs) by machine learning (ML) model using the original petrophysical data from an oil and gas sandstone reservoir in the Malay Basin and build a generalisable ML model. The ML framework is based on Cross Industry Standard Process for Data Mining (CRISP-DM) workflows and Exploratory Data Analysis (EDA) as an iterative cycle to analyse and tune the algorithms. First, appropriately address the composite data and its associated uncertainties through data pre-processing. Second, set the data splitting and evaluate the prediction model's through several regressions. Third, run an optimization algorithm to search for the best hyperparameters for the regressor to optimize the prediction. The ML method then captured the performance measure from the Coefficient of Determination (R^2) of 0.96 and 0.97 for Random Forest and Decision Tree Regression, respectively, and the lowest Root Mean Square Error (RMSE) value was recorded at 0.05, which indicates the excellent model with positive correlation. It is observed that the predicted Vp (DTC logs) and Vs (DTS logs) of the ML model produced good cross-validation to the original logs with a good performance measure of 1.0 for R^2 and 0.0 for RMSE. It can be concluded, based on the performance measure of each method, indicates the robustness of DTS log prediction using a quantitative measure of accuracy in scoring the predictions. It demonstrates the ML model's ability to generalize and predict shear logs on full field size.

ABSTRAK

Log sonik adalah parameter penting untuk keterangan takungan dan pengenalan cecair dan digunakan secara meluas dalam menentukan sifat batuan mekanikal untuk fizik batuan, tafsiran seismik kuantitatif dan aplikasi geomekanik. Selalunya terdapat kekurangan kecepatan gelombang ricih (Vs) data di telaga eksplorasi minyak dan gas yang relatif disebabkan oleh keadaan lubang bor yang buruk (pencucian), alat yang rosak, mengimbangi data dengan baik dan agak mahal. Ini bertujuan untuk memberikan penyelesaian untuk meramalkan gelombang mampatan (Vp) dan halaju gelombang ricih (Vs) dengan machine learning (ML) model menggunakan data petrofizik asli dari minyak dan takungan batu pasir gas di Malay Basin dan membina model ML umum. Kerangka ML didasarkan pada Cross Industry Standard Process for Data Mining (CRISP-DM) dan Exploratory Data Analysis (EDA) sebagai kitaran berulang untuk menganalisis dan menyesuaikan algoritma. Pertama, atasi data komposit dan ketidakpastian yang berkaitan dengan tepat melalui pemprosesan data. Kedua, tetapkan pemisahan data dan menilai model ramalan melalui beberapa regresi yang berbeza. Ketiga, jalankan algoritma pengoptimuman untuk mencari hiperparameter terbaik untuk regresor untuk mengoptimumkan ramalan. Kaedah ML kemudian menangkap ukuran prestasi dari Pekali Penentuan (R^2) 0.96 dan 0.97 untuk Random Forest dan Decision Tree Regression dan nilai Root Mean Square Error (RMSE) terendah dicatatkan pada 0.05 yang menunjukkan model yang baik dengan korelasi positif. Telah diperhatikan bahawa log Vp (DTC yang diramalkan) dan Vs (DTS log) model ML menghasilkan pengesahan silang yang baik ke log asal dengan ukuran prestasi yang baik 1.0 untuk R^2 dan 0.0 untuk RMSE. Ini dapat disimpulkan, berdasarkan ukuran prestasi setiap kaedah, menunjukkan ketahanan ramalan log DTS menggunakan ukuran ketepatan kuantitatif dalam menjaringkan ramalan. Ini menunjukkan kemampuan model ML untuk menggeneralisasi dan meramalkan log ricih pada ukuran medan penuh.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ML	-	Machine Learning
CRISP-DM	-	Cross Industry Standard Process for Data Mining
EDA	-	Exploratory Data Analysis
UTM	-	Universiti Teknologi Malaysia
ANN	-	Artificial Neural Network

LIST OF SYMBOLS

μ	-	Stiffness modulus
ρ	-	Rock density
φ	-	Formation porosity
ρ_i	-	Density of radioactive minerals
V_i	-	Minerals volume
A_i	-	Radioactive factor (radioactivity intensity of minerals)
ρ_b	-	Formation density

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Research Background

Sonic log data is an essential parameter for reservoir description and fluid identification and is extensively applied in determining mechanical rock properties for rock physics, quantitative seismic interpretation and geomechanics application. In actual application, there is frequently a paucity of shear wave velocity (Vs) information due to several factors, which only a few development wells have Vs log data, while old fields (brownfields) have been developed with longer development times have less to none of Vs log data. To date, when applying well seismic technology to fine reservoir parameter inversion based on seismic data in old fields, compressional wave velocity (Vp), Vs and density log are needed to establish formation strata modelling. The sonic well logs' unavailability is potentially due to the poor borehole conditions (washout), damaged tools and offset well data. Most offset well log data are not acquired with dipole sonic imager tools but with a borehole compensated logging tool, which limits the application of acoustic measurements to estimate the mechanical rock properties (Onalo et al., 2018). Empirical approaches are commonly used to estimate sonic velocity, and more recent research has demonstrated that neural networks can provide accurate predictions of formation sonic velocity. Nevertheless, these correlations cannot produce the desired results in different settings, even the most frequently used correlations such as those proposed by (Brocher, 2005; Eskandari et al., 2004; Greenberg & Castagna, 1992; Pickett, 1963). Meanwhile, because these correlations only consider the compressional wave velocity (Vp), they are only slightly capable of estimating shear wave velocity (Vs).

The synthetic sonic log offers more precise, reliable, and continuous indications of the formation's strength and mechanical rock properties. In addition, log data from the density and neutron logs can be utilised as a porosity indication for the formation. It has proven that the compressional wave velocity (Vp) and shear wave velocity (Vs) are widely used as quick, easy to use, and cost-effective means of determining the mechanical properties of the formation (Akhundi et al., 2014). However, shear wave logs are only available in a limited number of wells in an oil field due to the high cost of the log acquisition. Therefore, researchers are trying to estimate Vs from methods with acceptable accuracy like the empirical equation, as it should be more economical and only require the existing information. Integration with machine learning will create new statistical techniques to solve reservoir characteristics and geomechanical parameters that the empirical equation cannot achieve (Akhundi et al., 2014).

The uncertainty level of data in the reservoir exploration and development cycle is significantly high at the beginning of the exploration during the core sample data acquisition. However, the uncertainty will reduce when it reaches the appraisal and development level. The core data is quite diversified, containing images, waveforms, and numeric values with continuous and discrete depth indexes. Geological controls such as heterogeneity, engineering considerations such as operating conditions (drilling/logging), and physical sensors all contribute to variability in petrophysical data.

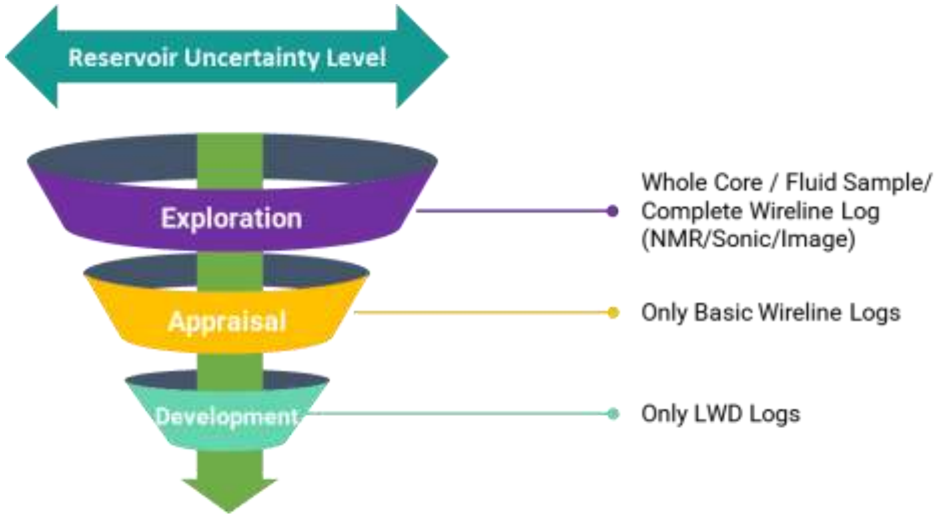

Figure 1.1 Petrophysical data acquisition for formation evaluation reduces as the reservoir uncertainty level decreases during the reservoir exploration and development cycle.

Petrophysical data is vital to visualize trends, patterns, distinctions, and clusters of the specific reservoir properties. The common practises are histograms, cross-plots, logging track displays, heatmaps and correlation graphs. The data has substantially impacted business decision-making.

Data quality is critical in ensuring the correct environmental-affected data, data reconstruction, and statistical correction and reconstruction processes for all well logs. Based on petrophysical principles, selecting and eliminating irrelevant data can flatten the prediction model's rate or margin of error, especially when dealing with multiple features and algorithms.

Compared to physics-based or data-driven models, petrophysical models are used to quantitatively determine various petrophysical parameters by processing physical measurements received from core or well logs. For example, water saturation calculation uses empirical Archie's model and its modifications through a combination of porosity estimations and resistivity logs (Archie, 1942). However, Archie's model makes various assumptions, including clay-free rocks, no major invasion, and the absence of sophisticated pore networks. Archie's model is insufficient for clay-rich, highly tortuous, and thinly laminated reservoirs (Worthington, 2000). When the assumptions of these models cannot be satisfied due to the complexity, heterogeneity, and multiscale nature of physical processes, these models become unsuitable for petrophysical interpretations and estimates.

The models required for petrophysical computations must be exceedingly non-linear and non-explicit. Such non-linearity is unexplainable by mechanistic, empirical, or phenomenological theories. As an alternative, constructing data-driven models using machine learning (ML) can improve the characterization of petrophysical processes and systems. When the deterministic physics-based model is unavailable, the data-driven models can give a computationally cheaper surrogate model to replace the costly physics-based model or provide a statistical approximation model based on the observations (Aifa, 2014).

The advent of big data and the more affordable computing hardware and software drive the recent resurgence of ML. Numerous endeavours in applying ML for petrophysical analysis, particularly in synthetic log generation or log prediction, such as shear sonic log prediction, receive immense attention. The recent resurgence of machine learning, due in part to the advent of big data and improved and more affordable computing hardware and software, has led to numerous endeavours in the application of machine learning for petrophysical analysis and particularly synthetic log generation or log prediction, and sonic log prediction has received particular attention (Akhundi et al., 2014; Eskandari & Rezaee, 2003; Rajabi et al., 2010).

When the prediction can be a solution to the absence of well logs data, value creation in terms of cost-saving is formulated and calculated. The cost of well activities such as rental wireline equipment, logistics, human resources, and lab analysis is calculated as determining factors for creating the business cases. The major cost reduction is estimated to be around RM 80k to RM 100k per curve, depending on the complexity of the well intervention during the phases, whether at the Exploration, Appraisal or Development, which would have an impact on the economic results of the specific project during project sanction. The focus of cost optimization efforts was typically done to improve project economics by re-examining the main contributors such as facility concepts, drilling or well's costs, hook-up and commissioning (HUC), which will be parked under CAPEX and operations and maintenance (O&M) is registered under OPEX as input to generate complete economic evaluations.

This paper aims to provide a solution to the machine learning (ML) model for the field engineers to predict the compressional wave (Vp) and shear wave velocities (Vs). Improvement in this research focuses on the prediction model involving machine learning algorithms and examines and evaluates the best-performing model using the actual dataset.

1.2 Problem Statement

Empirical regression and equation never consider factors like the fluid in pores, clay mineralogy, grain size, and bulk density of the rock, even though these factors affect the Vs, as recently highlighted in many works of literature. Similarly, factors like pore geometry, fracture orientation and intensity, bulk density, depth of burial, effective stress, and type of cementation are also not included. The following items are included in this section to address the issue of predicting shear wave:

(a) Vs is usually estimated from the analysis of core samples and using tools such as dipole sonic imager (DSI), which are expensive, relatively difficult to operate, time-consuming, and most of the DSI tools are not readily accessible in the wells.

(b) The empirical relationship excludes the various parameters affecting the shear wave velocity. Most empirical relations between Vp-Vs and petrophysical data are site-specific and only applicable to sandstone reservoirs.

(c) Scarcity of compressional travel time (DTC) and shear travel time (DTS) logs in all the wells drilled in a field due to financial or operational constraints. Under such circumstances, ML techniques can predict DTC and DTS logs to improve subsurface characteristics

1.3 Objective

The main objective of this thesis to code machine learning models to predict shear sonic log, Vs (DTS) using a real data from the Angsi field located in the Malay Basin. The specific objectives of this research are:

(a) To create a generalizable ML model specific to data-driven models using localized datasets (well logs).

(b) To cross validate the predicted sonic log with the actual sonic log based on the performance measure, coefficient of determination (R^2) and root mean squared error (RMSE).

1.4 Research Scope

This study was carried out based on the open dataset obtained from the Malaysia Petroleum Management (MPM) acts for and on behalf of PETRONAS. The scope is limited to the dataset of conventional logs identified as a pre-requisite to executing the research in the following components:

(a) Data gathering and data pre-processing:

i. The data included measured depth (MD), caliper log (CALI), gamma-ray log (GR), photoelectric factor (PEF), resistivity log (RT), bulk density (RHOB), neuron porosity (NPHI), compressional and shear sonic (DTC and DTS) as the main features.

ii. Differentiate and filtering of the datasets as requirement for quality control, model selection criterion, unbalanced data, checking the relevancy and redundancy

iii. Exploratory Data Analysis (EDA) - Plotting the cross plot, histograms, charts, heatmaps, and other forms for better data comprehension is readable and simplifies the visualization process

(b) Machine Learning (ML) modelling:

i. Using the cloud services software, Google Colaboratory (Colab) that capable of executing Python code interactively in a web browser. It can handle huge volumes of data that can be analyzed and has a great library ecosystem.

ii. Formulate and calculate ML algorithms like clustering, regression, and classification, which allows merging, filtering, training, and transforming data to the desired format.

(c) Performance Measure and Final Result.

i. The high prediction accuracy, quantified in terms of root mean squared error (RMSE), examines the coefficient of determination (R^2).

ii. Evaluating the final output of the predicted sonic log, Vs (DTS) generated from the ML model with the actual dataset as part of cross-validation.

1.5 Significant of Research

The interpretation of geological reservoir heterogeneity is one of the major reservoir uncertainties usually inferred from various data types: well logs, core data, seismic, production data analysis, and based on geological understanding to derive the conceptual interpretation. The data size often exceeds conventional dataset tools' capabilities to store, acquire, and analyse data. This is especially true given the data's non-linear nature and high complexity. As a result, models with higher levels of complexity and sophistication are vital in transforming raw data into scholarly output. Explanatory data analysis (EDA) identifies patterns and describes broad data. Predictive modeling relies on the interpretation of present variables to forecast future variables. The algorithms utilize past and current information in the dataset to uncover hidden trends for hypothesising descriptive and predictive models with decent generalizability.

Therefore, this thesis provides the computing ML solution to estimate the synthesise sonic logs from wells with missing well log data and erroneous data due to faulty or damaged tools. Thus, it offers a considerably cheaper alternative to running actual sonic logging tools like borehole compensated and dipole sonic logging tools. It is also an option to use empirical correlation and offset data to determine the sonic log for calculating mechanical rock properties.

Ultimately, the research aims to improve the accuracy and efficiency of Upstream Exploration by using a reliable ML model to predict the sonic log as an essential input for geophysical and geo-mechanical studies in identifying rock strength and understanding the stress regime.

REFERENCES

Akhundi, H., Ghafoori, M., & Lashkaripour, G.-R. (2014). Prediction of Shear Wave Velocity Using Artificial Neural Network Technique, Multiple Regression and Petrophysical Data: A Case Study in Asmari Reservoir (SW Iran). *Open Journal of Geology*, *04*(07), 303–313.

Ameen, M. S., Smart, B. G. D., Somerville, J. M., Hammilton, S., & Naji, N. A. (2009). Predicting rock mechanical properties of carbonates from wireline logs (A case study: Arab-D reservoir, Ghawar field, Saudi Arabia). *Marine and Petroleum Geology*, *26*(4), 430–444.

Asoodeh, M., & Bagheripour, P. (2012). Prediction of compressional, shear, and Stoneley wave velocities from conventional well log data using a committee machine with intelligent systems. *Rock Mechanics and Rock Engineering*, *45*(1), 45–63.

Asquith, G. B., Gibson, C. R., Henderson, Steven., & Hurley, N. F. (2004). *Basic well log analysis*. American Association of Petroleum Geologists.

Azeem, T., Chun, W. Y., Monalisa, Khalid, P., Qing, L. X., Ehsan, M. I., Munawar, M. J., & Wei, X. (2017). An integrated petrophysical and rock physics analysis to improve reservoir characterization of Cretaceous sand intervals in Middle Indus Basin, Pakistan. *Journal of Geophysics and Engineering*, *14*(2), 212–225.

Breiman, L. (2001). *Random Forests* (Vol. 45).

Brocher, T. M. (2005). Empirical relations between elastic wavespeeds and density in the Earth's crust. *Bulletin of the Seismological Society of America*, *95*(6), 2081–2092.

Bukar, I., Adamu, M. B., & Hassan, U. (2019). *SPE-198764-MS A Machine Learning Approach to Shear Sonic Log Prediction*. http://onepetro.org/SPENAIC/proceedings-pdf/19NAIC/2-19NAIC/D023S026R001/1183051/spe-198764-ms.pdf/1

Ceryan, N. (2014). Application of support vector machines and relevance vector machines in predicting uniaxial compressive strength of volcanic rocks. *Journal of African Earth Sciences*, *100*, 634–644.

Cook, R. Dennis., & Weisberg, S. (1999). *Applied regression including computing and graphics*. Wiley.

Eskandari et. al. (2004). *Application of Multiple Regression and Artificial Neural Network Techniques to Predict Shear Wave Ve _ CSEG RECORDER*.

Eskandari, H., & Rezaee, R. (2003). *Shear Wave Velocity Estimation Utilizing Wireline Logs for a Carbonate Reservoir, South-West Iran*. https://www.researchgate.net/publication/266465565

Gaymard, R., & Poupon, A. (n.d.). Response neutron and formation density logs in hydrocarbon bearing formations effect of hydrocarbons in the neutron log. http://onepetro.org/petrophysics/article-pdf/2197605/spwla-1968-vixn5a1.pdf/1

Greenberg', M. L., & Castagna3, J. P. (1992). Shear-wave velocity estimation in porous rock : theoretical formulation, preliminary verification and applications. In *Geophysical Prospecting* (Vol. 40).

Hadi, F., & Nygaard, R. (2018). *Shear Wave Prediction in Carbonate Reservoirs: Can Artificial Neural Network Outperform Regression Analysis?* https://www.researchgate.net/publication/327700812

Han, D.-H., Nur, A., & Morqan, D. (1986). Effects of porosity and clay content on wave velocities in sandstones. In *GEOPHYSICS* (Vol. 51, Issue 11). http://library.seg.org/

Holdaway, K. R. (2009). *SPE/EAGE Reservoir Characterization and Simulation Conference*.

Hossain, T. M., Watada, J., Aziz, I. A., Hermana, M., Meraj, S. T., & Sakai, H. (2021). Lithology prediction using well logs: A granular computing approach. *International Journal of Innovative Computing, Information and Control*, *17*(1), 225–244.

Maleki, S., Moradzadeh, A., Riabi, R. G., Gholami, R., & Sadeghzadeh, F. (2014). Prediction of shear wave velocity using empirical correlations and artificial intelligence methods. *NRIAG Journal of Astronomy and Geophysics*, *3*(1), 70–81.

Miah, M. I. (2020). Predictive models and feature ranking in reservoir geomechanics: A critical review and research guidelines. In *Journal of Natural Gas Science and Engineering* (Vol. 82). Elsevier B.V.

Miah, M. I., Ahmed, S., & Zendehboudi, S. (2021). Model development for shear sonic velocity using geophysical log data: Sensitivity analysis and statistical assessment. *Journal of Natural Gas Science and Engineering*, *88*.

Oloruntobi, O., & Butt, S. (2020). The shear-wave velocity prediction for sedimentary rocks. *Journal of Natural Gas Science and Engineering*, *76*.

Onalo, D., Adedigba, S., Khan, F., James, L. A., & Butt, S. (2018). Data driven model for sonic well log prediction. *Journal of Petroleum Science and Engineering*, *170*, 1022–1037. https://doi.org/10.1016/j.petrol.2018.06.072

Pickett, G. R. (1963). *Acoustic Character Logs and Their Applications in Formation Evaluation*. http://onepetro.org/JPT/article-pdf/15/06/659/2213525/spe-452-pa.pdf/1

Rajabi, M., Bohloli, B., & Gholampour Ahangar, E. (2010). Intelligent approaches for prediction of compressional, shear and Stoneley wave velocities from conventional well log data: A case study from the Sarvak carbonate reservoir in the Abadan Plain (Southwestern Iran). *Computers and Geosciences*, *36*(5), 647–664

Raymaekers, J., & Rousseeuw, P. J. (2021). Transforming variables to central normality. *Machine Learning*.

Shahab D. Mohaghegh. (2005). *Recent Developments in Application of Artificial Intelligence in Petroleum Engineering*.

Tosaya, C. A. (1983). *Acoustical properties of clay-bearing rocks*.

Xu, C. (2019). *SPE-195068-MS When Petrophysics Meets Big Data: What can Machine Do?*

Yeo, I., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. In *Biometrika* (Vol. 87, Issue 4). http://biomet.oxfordjournals.org/

Zoveidavianpoor, M., Samsuri, A., & Shadizadeh, S. R. (2013). Adaptive neuro fuzzy inference system for compressional wave velocity prediction in a carbonate reservoir. *Journal of Applied Geophysics*, *89*, 96–107.