ISSN: 0067-2904

Volve Oil Field S-Wave Log Data Prediction Using GBR and MLPR

Amany G. Fadhil [1]*, Hana M. Ali [1], Zainab A. Khalaf [1], Musa Ahmed [2], Semaa H. Ahmed [2]

[1]Department of Mathematics, College of Sciences, University of Basrah, Basrah, Iraq
[2]Department of Petroleum Engineering, College of Engineering, University of Houston, Texas, USA

Received: 11/12/2022 Accepted: 22/4/2023 Published: 30/4/2024

Abstract

The shear wave sonic (S-wave) log data is essential for identifying the reservoir's geomechanical properties, which is an important factor for the drilling, completion, and optimization processes, where obtaining S-wave requires capital investment and reduces the cost. In this paper, we used the multi-layer perceptron (MLPR) and the gradient boosting regression (GBR) to predict the missing S-Wave log data for the Volve Oil Field in the North Sea. Prescriptive and predictive analysis were carried out in series to achieve a high blind data accuracy rate of 0.943 and 0.982 for the multi-layer perceptron regression and the gradient boosting regression, respectively. It was observed that the gradient-boosting algorithm achieved higher accuracy than the MLPR algorithm for the limited dataset. It was also found that hierarchical clustering can reveal information regarding the feature's importance similar to the relevant AI tools, which makes hierarchical clustering a faster tool in eliminating the nonimportant inputs from the dataset while the use of artificial intelligence tools showed a significant effect in predicting the missing values of the sonic wave log and the neutron porosity log in an efficient way by selecting the relevant important features.

Keywords: Artificial Neural Networks, Decision Trees, Gradient Boosting, hierarchical clustering, shear wave sonic.

التنبؤ بسجل بيانات الموجات الصوتية لحقل فولفو النفطي باستعمال طريقة الإدراك متعدد الطبقات والانحدار التدريجي

اماني جمال فاضل[1]*, هناء مرتضى علي[1], زينب علي خلف[1], موسى احمد[2], سيماء حاتم احمد [2]

[1]قسم الرياضيات، كلية العلوم, جامعة البصرة، البصرة, العراق

[2]قسم هندسة البترول, كلية الهندسة, جامعة هيوستن, تكساس, الولايات المتحدة الامريكية

الخلاصة

تعد بيانات السجل الصوتي لموجة القص ضرورية لتحديد الخصائص الجيوميكانيكية للمكمن النفطي، وهو عامل مهم لعملية الحفر والإكمال والتحسين، حيث يتطلب الحصول على السجل الصوتي لموجة القص

* Email: Sci.lect.004@avicenna.uobasrah.edu.iq

استثماراً رأسمالياً ولتقليل الكلفة، استعملنا في هذا البحث انحدار الإدراك متعدد الطبقات وانحدار تعزيز التدرج للتنبؤ ببيانات السجل الصوتي لموجة القص المفقودة لحقل نفط فولفي في بحر الشمال. تم إجراء التحليل الوصفي والتنبؤي على التوالي لتحقيق معدل دقة بيانات أعمى عالي، حيث بلغ 0.943 و0.982 لانحدار الإدراك متعدد الطبقات وانحدار تعزيز التدرج، على التوالي. لوحظ أن خوارزمية تعزيز التدرج حققت دقة أعلى من خوارزمية انحدار الادراك متعدد الطبقات لمجموعة البيانات المحدودة. كما وجد أن التجمع الهرمي يمكن أن يكشف عن معلومات تتعلق بأهمية الميزة مماثلة لأدوات الذكاء الاصطناعي ذات الصلة، مما يجعل التجمع الهرمي أداة أسرع في إزالة المدخلات غير المهمة من مجموعة البيانات، بينما أظهر استعمال أدوات الذكاء الاصطناعي تأثيرًا كبيرًا في التنبؤ بالقيم المفقودة لسجل الموجة الصوتية وسجل مسامية النيوترون بطريقة فعالة من خلال اختيار الصفات المهمة ذات الصلة.

1. Introduction

Fossil fuel is considered one of the most expensive commodities, especially with the increase in fossil fuel prices and the global demand for a clean energy transition. Therefore, utilizing alternatives associated with the fourth paradigm shift of data-driven science helped offer the fossil industry a relatively cheaper and faster decision-making tool in many sectors. Artificial intelligence (AI), which is data-driven, has been widely used for data prediction in different sectors [1–4], and it has many applications within the oil and gas industry [5–12]. Shear Wave Sonic (S-wave) log data is essential for identifying the reservoir geomechanical properties [13], which is an important factor for the drilling, completion, and optimization processes, where obtaining an S-wave log requires capital investment. At the reservoir's time, data quality depends on the wellbore conditions, where unconsolidated formations and shale may influence data availability and accuracy due to S-wave refraction from the tool. Therefore, as the value of obtaining new information is high, it is essential to employ alternatives that model past data to generate synthetic S-wave logs, such as artificial intelligence (AI) modeling.

Lately, different authors have investigated different AI methods for predicting the S-Wave log for various fields [14–16]. For example, [15] in 2019 used empirical correlation and two AI methods, Support Vector Regression (SVR) and Back-Propagation Neural Network, to predict the S-Wave log. Their study concluded that the SVR method was relatively better. On the other hand, [9] used regional logging data with SVR, eXtreme Gradient Boosting (XGBoost), random forest, linear regression, and ANN to predict the S-Wave log. The outcomes demonstrate that XGBoost outperforms the other models with R^2 of 0.994 and 0.964 for training and testing, respectively. Similarly, [17] proposed a new approach to predict the S-Wave log and the compressional wave sonic (P-Wave) log in the Volve oil field. They used different AI methods such as artificial neural networks (ANN), adaptive neuro-fuzzy inference systems (ANFIS), and least square support vector machines (LSSVM). Their model demonstrated excellent accuracy in predicting the S-wave log of 0.99 R^2 with the LSSVM outperforming the other methods.

In this paper, a synthetic shear wave sonic (S-Wave) log is predicted using two AI methods: artificial neural networks (ANN) and gradient boosting regression (GBR). We exploited the largest open-source oilfield database, the Volve field, in predicting the S-Wave log.

2. Methodology

In this paper, a prediction system is proposed to predict a missing wave sonic log based on the predefined value of a known wave sonic log. The proposed system consists of four main

phases. These phases are raw dataset, processed dataset, feature selection, and model selection, as shown in Figure 1.

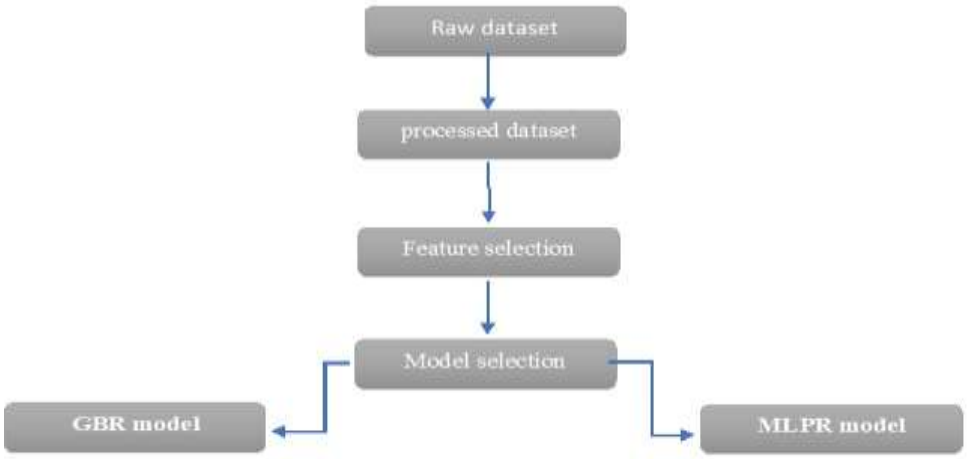

Figure 1: flowchart of methodology.

2.1 Raw dataset

Volve is an oil field located in the southern part of the North Sea, 200 kilometers west of Stavanger. Volve Field was developed with 24 wells, and a total production of 63 million barrels of oil was produced during the eight-year field lifetime starting in 2008. In 2018, Equinor published vast amounts of public data, ranging from well data to the entire geological model of the Volve field for study, research, and development [18]. This research used the available well log data to predict the S-wave log.

2.2 Processed dataset

Initially, the total number of wells with well log data was 21, with 19,342,000 bytes of data. After dropping the missing cells and keeping the available wells with S-wave logs with the basic well log sets, the final well count was 5 with a 481,370 data size that was used for training and testing. This model intends to obtain high accuracy to fill the missing S-wave data count of 2,161 within the trained five wells. The available well logs after data cleaning were: measured depth (MD), true vertical depth (TVD), neutron porosity (NPHI), bulk density (RHOB), density correction (DRHO), gamma ray (GR), resistivity (RT), photoelectric factor (PEF), calliper (CALI), p-wave (DT), and shear wave (DTS). Figure 1 shows an example of 2 out of 5 wells (wells 15/9-F-11 T2 and 15/9-F-1 B) with the logs used for building the S-wave predictive model. Figure 2 track 1 contains the GR, which is the total formation radioactivity content, measured in API. Track 2 contains RT, which represents the formation's true resistivity, measured in ohms per meter. Track 3 contains RHOB and NPHI, which are the formation bulk density, measured in grams per cubic centimeter, and the formation porosity, measured in v/v, respectively. Track 4 contains the DRHO, which is the RHOB correction factor, measured in grams per cubic centimeter. Track 5 includes DT and DTS, the P-Wave log, measured in microseconds per foot, and the S-Wave log travel time, respectively.

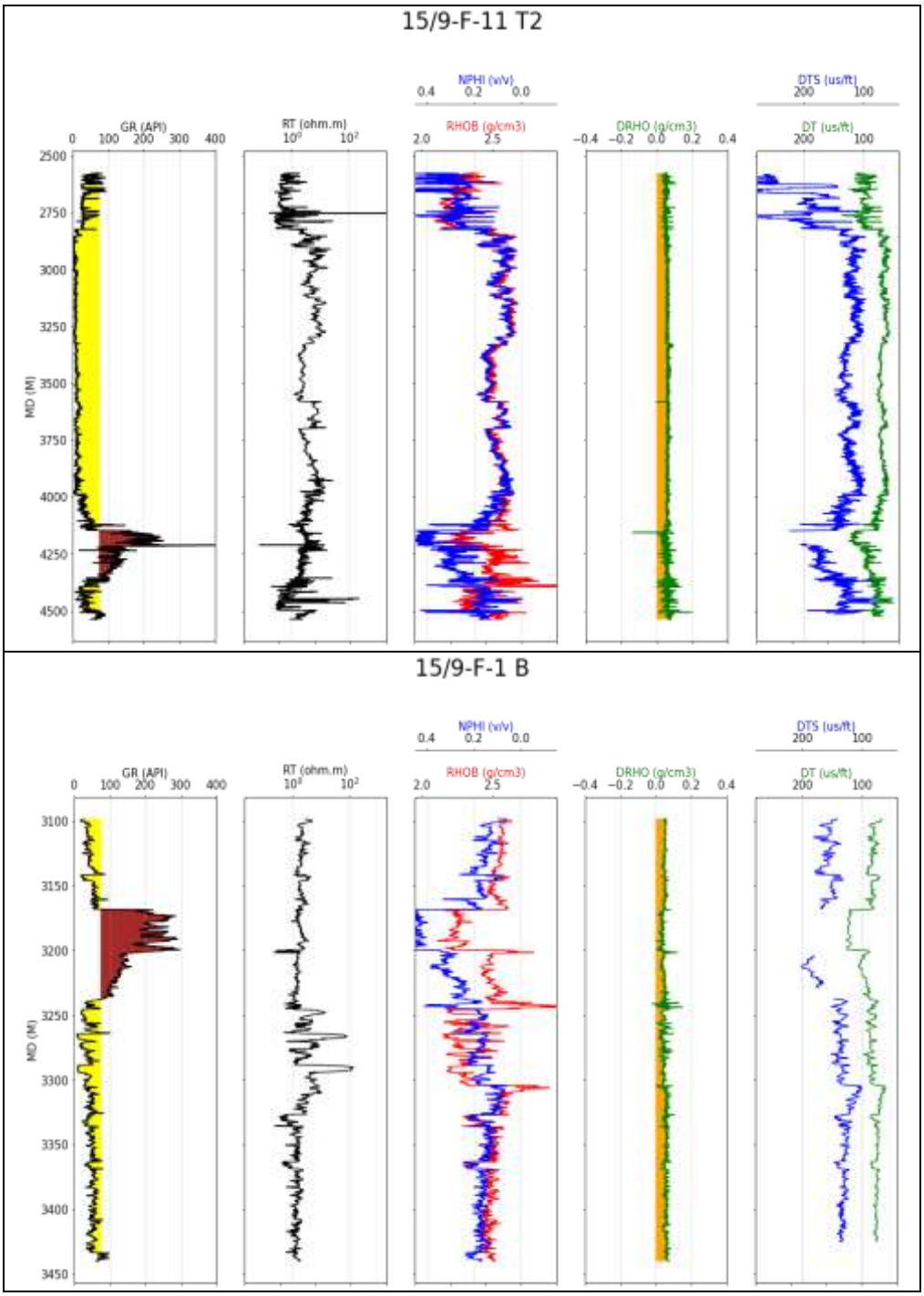

Figure 2: well logs example for two wells (15/9-F-11 T2 and 15/9-F-1 B) used in building the predictive model.

2.3 Feature selection

Descriptive analytics was performed to find trends within the input data and avoid unnecessary data. Hierarchical clustering analysis was performed to approximate input/output data relationships. It analyzes large amounts of data by grouping them according to some measure of similarity and finding a way to assess how they differ and overlap [19] and [20]. Hierarchical clustering is divided into two types: agglomerative, where each observation begins in its own cluster, and as one climbs the hierarchy, pairs of clusters are combined ([20], [22], and [23]); and divisive, in which each observation begins in its own cluster, and as

one descends the hierarchy, splits are carried out again ([23], [25], and [26]). Herein, we used Euclidean distance ([27], [28], and [29]) as a similarity metric to measure the dissimilarity between the sets of data.

Hierarchical clustering calculation was performed, and results are shown in Figure 5(a) on a type of tree diagram called a dendrogram demonstrating the hierarchical grouping. It was concluded from the figure that DT (P-wave log) and NPHI share more similarities with DTS (S-wave log), whereas GR comes in second in similarity, while RT and RHOB share minor similarities with DTS. The mathematical grouping is logical, as DTS is not affected by the rock fluids themselves and is a function of the rock matrix, whereas RT and RHOB are influenced by both rock fluids and matrix (bulk).

A different approach was also used to find patterns by using the Random Forest (RF) Feature Importance Algorithm of Mean Decrease Impurity (MDI), which calculates the total reduction in loss or impurity contributed by all splits for a given feature [30]. The importance of the feature classifies the features according to their impact on the model because the selected feature is the one that will have the greatest impact on it. They are useful in understanding the data, interpreting the model, and improving the model's performance by making it simpler and faster because it eliminates lower scores.

There are several ways to calculate the importance of the feature, including mean decrease impurity (MDI), which calculates the impurities of the node and reduces them according to the number of samples. MDI calculates the importance of each feature in proportion to the number of samples that have been divided, leading to improvements in the model and a better understanding of the problem through feature selection. While The feature of permutation importance (PIMP) calculates the importance of switching the predictive value of the feature by knowing the increase or decrease in the error while switching feature values (i.e., measuring the prediction error for the model when the feature is not available). The PIMP does not depend on the model but provides insight into the model's behavior through the influence of each feature on the model's prediction [31]. RF feature importance results are presented in Figure 5 (b), where RF MDI and PIMP estimated that NPHI and DT had the highest influence while RT had the least.

2.4 Model selection

In this step, we utilized 2 different AI approaches to build the S-wave predictive model: the multilayer perceptron regression (MLP) and the gradient boosting regression (GB). Both models followed different approaches in terms of data pre-processing and hyperparameter selection techniques, while both shared similar data partitioning selection, where 30% of the data was set aside to validate the model and 70% of the training data was used for the training purpose. The random seed number was 2. It is important to note that the authors tested different random seed partitionings to cover a variety of features in order to improve the predictive model.

2.4.1 MLPR model

Multilayer Perceptron (MLP) is a feedforward ANN made up of successive layers that exchange and interconnect data via connections between neurons, which are expressed by an adaptative weight. The multilayer network's structure includes an input layer with several neurons equivalent to the number of input data attributes (MD, TVD, NPHI, RHOB, DRHO, GR, RT, PEF, CALI, and DT) and an output layer where the number of neurons is equal to the number of outputs (DTS) to predict, whereas the calculations are carried through the

hidden layers [32] as shown in Figure 2, where each neuron is connected to the adjacent-layer neurons, and the weight defines the degree of the connection, where f_Y, f_K, and f_J are the activation functions (ReLu) used for the neurons, and w_{Xm}, w_{Jo} and w_{Kq} are the weights. The coefficients b_{Xn}, b_{Jp}, and b_{Kr} are the biases assigned to shift the activation function by adding a constant within the neurons in the hidden layer (s).

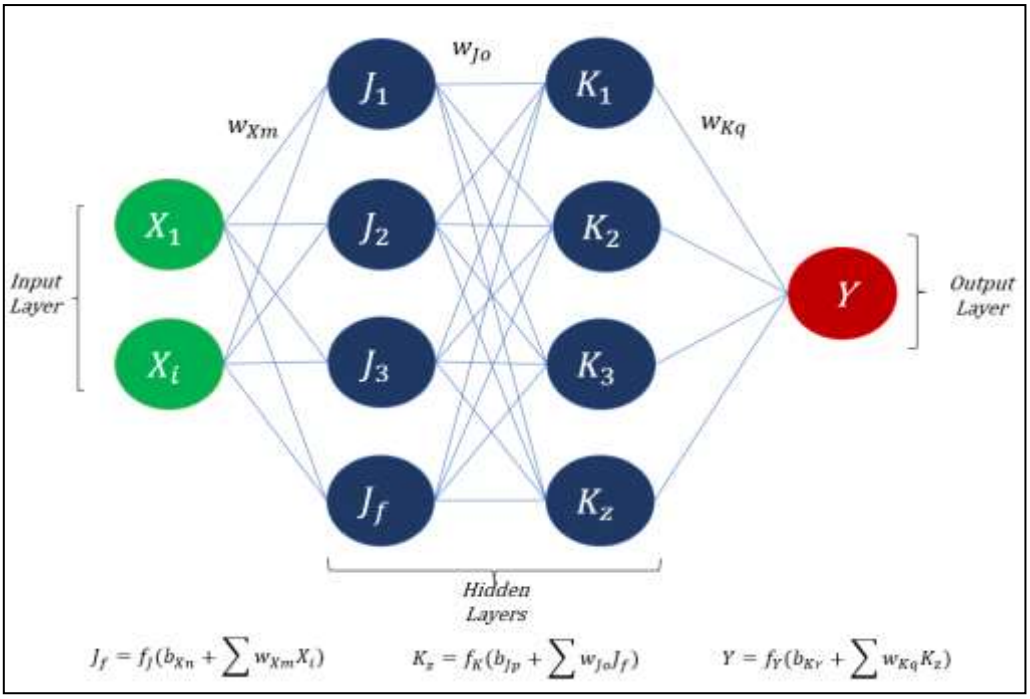

Figure 3: Simple MLP topology showing the neurons within the input layer, hidden layer, and the output layers.

MLP training was accomplished in two steps: data transformation and hyperparameter selection. In the data transformation section, data was normalized using Yeo Jonson power transformation, whereas the power parameter lambda (λ) is changed until the best approximation of a Gaussian likely distribution is reached [33]:

$$\psi(\lambda, y)\begin{cases}((y+1)^{\lambda}-1)/\lambda & if\ \lambda \neq 0, y \geq 0\\ log(y+1) & if\ \lambda = 0, y \geq 0\\ -[(-y+1)^{2-\lambda}-1)]\ /\ (2-\lambda) & if\ \lambda \neq 2, y < 0\\ -log(-y+1) & if\ \lambda = 2, y\ < 0\end{cases} \quad (1)$$

Within the hyperparameter selection, a loop of five layers of different hidden neuron combination lists was tested to include 35, 60, and 85 hidden neurons with Adam solver and ReLu activation functions. An early stopping feature with a 10% validation dataset was used to stop the training process when overfitting is expected. where the maximum number of iterations is set to 3000.

2.4.2 GBR model

Gradient Boosting Regression (GBR) has the advantage of better accuracy, higher flexibility, and fewer pre-processing requirements when compared with other regression methods. GB computes the variance between the present forecast and the identified true target value. This variation is referred to as the residual. After that, a weak model that maps features to that residual is trained using gradient boosting regression [34], as illustrated in Figure 4.

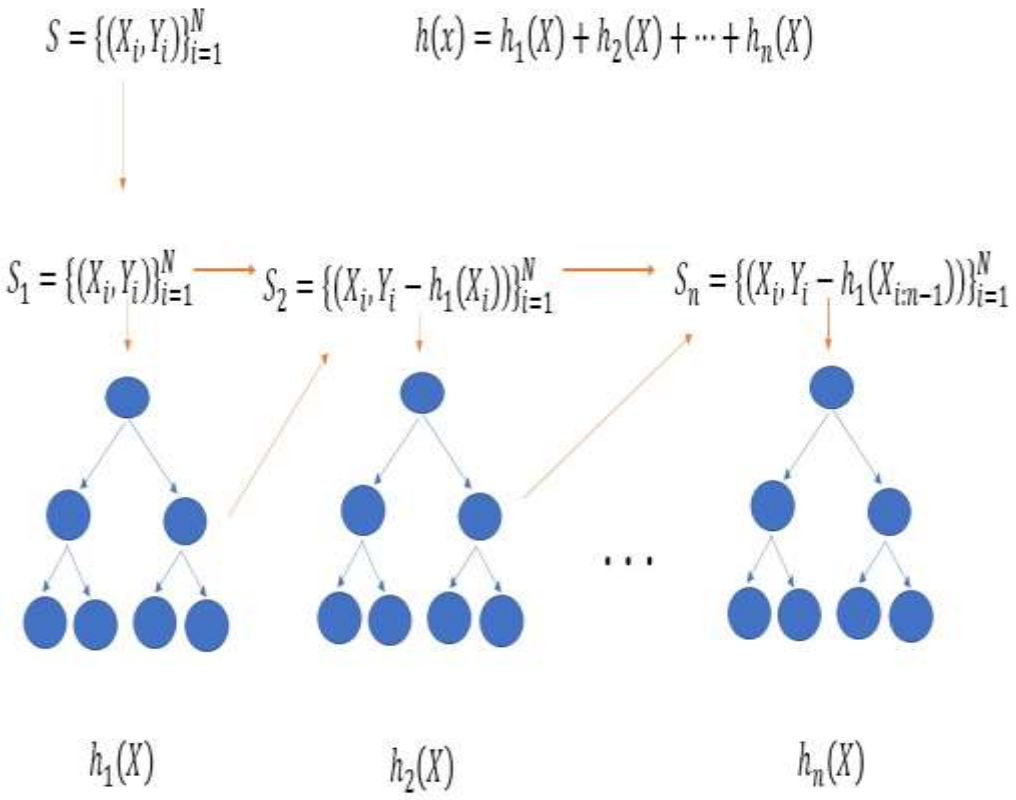

Figure 4: simple version gradient boosting regression model showing.

As GB doesn't require data normalization, it became advantageous over the suggested MLPR model; therefore, only hyperparameter selection was needed. The grid search technique was used with a cross-validation of 7. The tested parameters are the maximum depth and the number of estimators. 400 and 500 estimators were tested, along with 10, 30, 50, 70, 90, and 100 maximum GB depths.

It is also important to note that data quality control is essential to ensure data input accuracy for an accurate prediction for the missing S-wave log, which involves the wellbore condition and the depth of investigation of the input data to cover the targeted wellbore.

3. Results and Discussion

In this section, the results of the proposed system are shown and discussed based on feature selection results and model selection outcomes.

3.1 The feature selection results

The carried descriptive analysis provided insight for improvement within the input data if higher accuracy results were not achieved with the initial run by dropping the least influential attributes for better results.

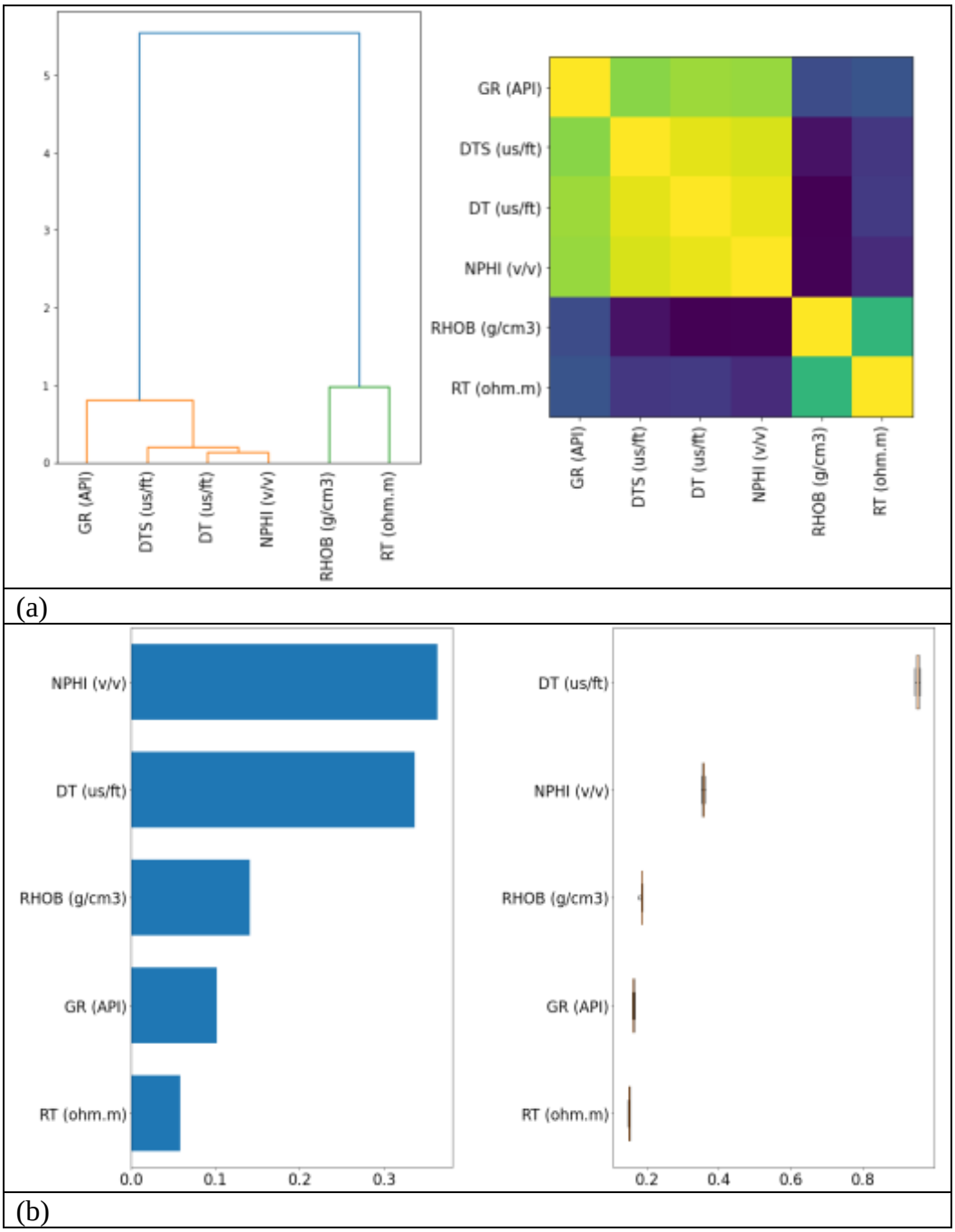

Figure 5: descriptive analytics with (a) Hierarchical clustering and (b) Random Forest Feature Importance (left) and Permutation (right).

3.2 The model selection results

The best result yielded 5 layers with 85 hidden neurons, and the highest $\mathbf{R^2}$ values achieved for training and testing were 0.96 and 0.94, respectively. While using a single hidden layer with up to 100 hidden neurons didn't achieve good results due to the high complexity of the data and the fact that shear wave sonic is more prone to distortion as compared to compressional wave sonic. The best results were achieved at 500 estimators and 10 maximum depths, with $\mathbf{R^2}$ values of 0.99 and 0.98 for training and testing, respectively. The final predictive model results are shown in Figure 6, where GB achieved better results with less scattered testing data while the MLP predictive model had relatively more noise. Both methods' computation times were large in our runs due to the large GB maximum depths and the large number of layers and neurons in the MLP; nevertheless, the MLP was

faster than the GB. It is due to the fact that multi-layer perceptron regression trains all the neurons in the network at once, whereas gradient-boosting regression trains numerous weak learners sequentially. This process will also affect the memory usage, as it is higher for the gradient-boosting regression than for the multi-layer perceptron regression. This is so that gradient-boosting regression can keep all of the weak learners' predictions during training. Figure 7 shows an example of the predictive model S-wave in blue and the predicted missing S-wave values in red, where both MLP and GB showed almost similar predictive patterns.

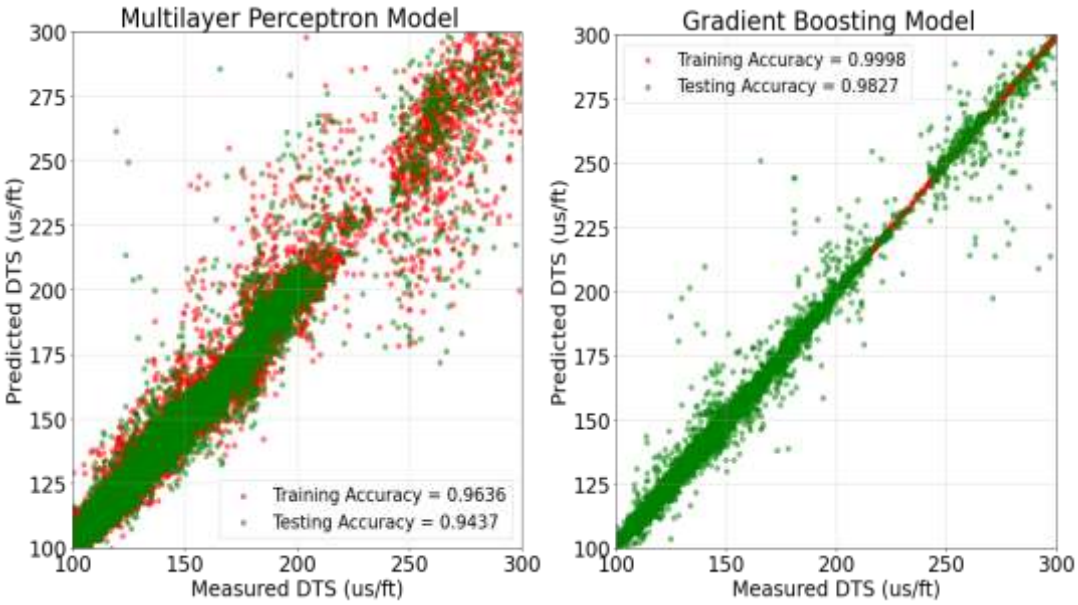

Figure 6: 5 wells' S-Wave predictive model accuracy rate for the 70% training and the 30% testing data where the red dots represent the training data and the green dots represent the blind testing data for MLP (left) and GB (right)**.**

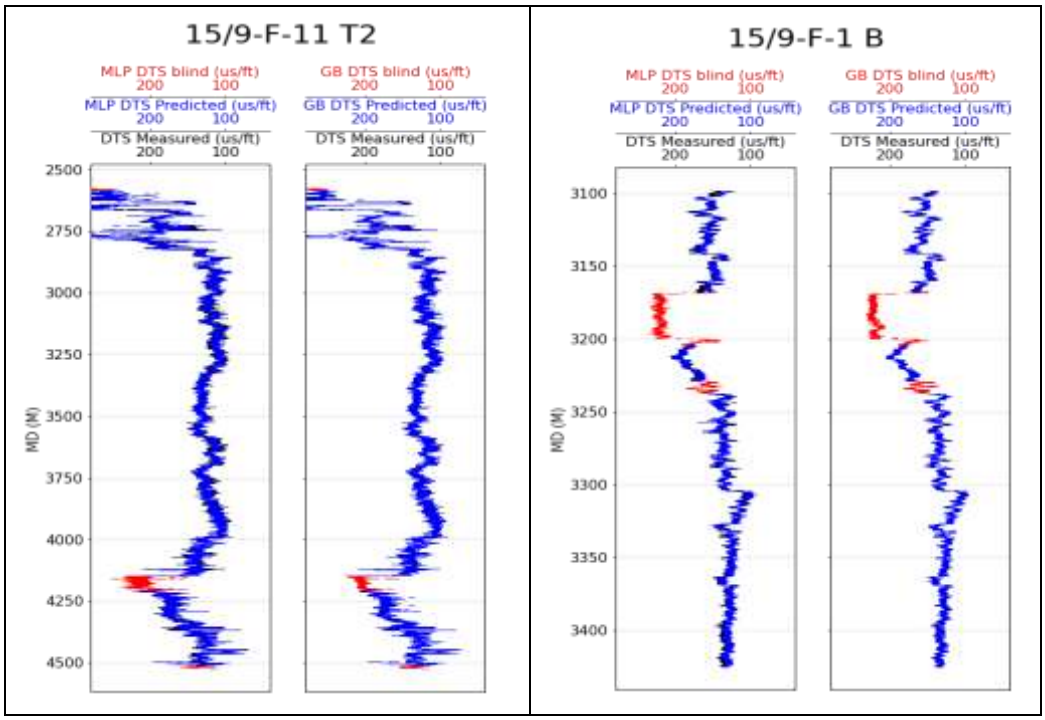

Figure 7: wells (15/9-F-11 T2 and 15/9-F-1 B) measured and predicted S-wave logs where the black curve represents the actual S-wave log, the blue curve refers to the predicted S-wave logs and the red curve is the predicted missing values.

4. Conclusion

S-wave log descriptive and predictive analytics were performed for the Volve field by training five wells with ten inputs. The descriptive analysis showed a great influence of DT and NPHI on both hierarchical clustering and random forest feature importance. However, for RF feature importance, the tool suggests that RT is the least influential parameter.

The GB regression was advantageous over the MLP regression with 0.98 $\mathbf{R^2}$ accuracy for the testing data, while the MLP S-wave predictive model showed some relative noise where it was impossible to achieve higher accuracy for training and testing with a single layer regardless of the number of hidden neurons.

5. Acknowledgements

Many thanks to Equinor and the former Volve license partners for making Volve field data available for public use.

References

[1] S. F. Raheem and M. Alabbas, "Dynamic Artificial Bee Colony Algorithm with Hybrid Initialization Method," *Informatica*, vol. 45, pp. 103–114, 2021.

[2] J. K. Al Saedi, R. D. Alalawy, and N. M. Hasan, "Determination the Concentration Elements of Cultivation Media (Peat Moss, Perlite and Hormone) Using X-ray Fluorescence Technique", *Iraqi Journal of Science*, vol. 59, no. 4A, pp. 1769–1777, Oct. 2018.

[3] L. M. Alhelfi and H. M. Ali, "Using Persistence Barcode to Show the Impact of Data Complexity on the Neural Network Architecture," *Iraqi Journal of Science*, vol. 63, no. 5, pp. 2262–2278, 2022, Doi: 10.24996/ijs.2022.63.5.37.

[4] H. S. Almalikee and Fahad M. Al-Najm, "Optimization of Deviated and Horizontal Wells Trajectories and Profiles in Rumaila Oilfield," *J. Pet. Res. Stud. Optim.*, No.27, p. 6, 2020.

[5] M. Farahnakian, M. R. Razfar, M. Moghri, and M. Asadnia, "The selection of milling parameters by the PSO-based neural network modeling method," *International Journal of Advanced Manufacturing Technology*, vol. 57, no. 1–4. pp. 49–60, 2011. Doi: 10.1007/s00170-011-3262-1.

[6] A. Hassan, M. Mahmoud, A. Al-Majed, and A. Abdulraheem, "A New Technique to Quantify the Productivity of Complex Wells Using Artificial Intelligence Tools," *International Petroleum Technology Conference IPTC, King Fahd University of Petroleum & Minerals*, 2020.

[7] M. A. U. Naser, "Prediction prices of Basra light oil using artificial neural networks," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 3, pp. 2682–2689, 2020, Doi: DOI: 10.11591/ijece.v10i3.

[8] H. Li, H. Yu, N. Cao, H. Tian, and S. Cheng, "Applications of Artificial Intelligence in Oil and Gas Development," *Archives of Computational Methods in Engineering*, vol. 28, no. 3. pp. 937–949, 2021. Doi: 10.1007/s11831-020-09402-8.

[9] Shanshan Liu, Yipeng Zhao, Zhiming Wang, "Artificial Intelligence Method for Shear Wave Travel Time Prediction considering Reservoir Geological Continuity," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5520428, 18 pages, 2021. https://doi.org/10.1155/2021/5520428

[10] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist, and H. Oza, "Application of machine learning and artificial intelligence in oil and gas industry," *Petroleum Research*, vol. 6, no. 4. pp. 379–391, 2021. Doi: 10.1016/j.ptlrs.2021.05.009.

[11] L. M. Alhelfi, H. M. Ali, and S. H. Ahmed, "P-Wave Sonic Log Predictive Modeling with Optimal Artificial Neural Networks Topology," *Journal of Al-Qadisiyah for Computer Science and Mathematics*, vol. 13, no. 3, pp. 142-154, 2021. Doi: /doi.org/10.29304/jqcm.

[12] M. A, T. Zeeshan, M. Mohamed, A. E. Radwan, A. Abdulraheem, and M. O. Abouelresh, "Data-driven machine learning approach to predict mineralogy of organic-rich shales: An example from Qusaiba Shale, Rub' al Khali Basin, Saudi Arabia, Marine and Petroleum Geology," *Mar. Pet. Geol.*, vol. 137, no. 0264–8172, 2022, Doi: doi.org/10.1016/j.marpetgeo.2021.105495.

[13] C. Naville, N. Cuenot, A. Tselentis, K. Kazemi, S. Serbutoviez, and J. Bruneau, "S-Wave Birefringence Variations From Stress, Pore Pressure?," *55th U.S. Rock Mech. Symp.*, p. ARMA-2021-1794, 2021.

[14] E. Salaheldin, T. Zeeshan, M. Mohamed, M. Ibrahim, and A. Abdulazeez, "Development of New Mathematical Model for Compressional and Shear Sonic Times from Wireline Log Data Using

Artificial Intelligence Neural Networks (White Box) Salaheldin," *Arab. J. Sci. Eng.*, vol. 43, pp. 6375–6389 ,2018. Doi: doi.org/10.1007/s13369-018-3094-5.

[15] S. Maleki, A. Moradzadeh, R. G. Riabi, R. Gholami, and F. Sadeghzadeh, "Prediction of shear wave velocity using empirical correlations and artificial intelligence methods," *NRIAG J. Astron. Geophys.*, vol. 3, no. 1, pp. 70–81, 2014, Doi: 10.1016/j.nrjag.2014.05.001.

[16] V. Roy, A. Gupta, R. Agrawal, N. Kumar, and A. Saxena, "Augmenting and eliminating the use of sonic logs using artificial intelligence: A comparative evaluation," *Geophys. Prospect.*, 2022, Doi: https://doi.org/10.1111/1365-2478.13213.

[17] T. Olayiwola and O. A. Sanuade, "A data-driven approach to predict compressional and shear wave velocities in reservoir rocks," *Petroleum*, vol. 7, no. 2, pp. 199–208, 2021, Doi: 10.1016/j.petlm.2020.07.008.

[18] J. Nilsson, "Disclosing all Volve data." Wooldridge, J. M., 2016, Introductory econometrics: A modern approach: Nelson Education," 2018. [Online]. Available: https://www.equinor.com/en/news/14jun2018-disclosing-volve-data.html

[19] S. C. Johnson, "Hierarchical clustering schemes," Psychometrika, vol. 32, pp. 241–254 (3684, 1967.

[20] F. Nielsen, "Undergraduate Topics in Computer Science Introduction to HPC with MPI for Data Science," *Springer*, 2016, pp. 195–211. Doi: DOI 10.1007/978-3-319-21903-5.

[21] I. Davidson and S. S. Ravi, "Agglomerative Hierarchical Clustering with Constraints: Theoretical and Empirical Results," *Eur. Conf. Princ. Data Min. Knowl. Discov.*, pp. 59–70, 2005.

[22] W. H. E. Day and H. Edelsbrunner, "Efficient Algorithms for Agglomerative Hierarchical Clustering Methods William," *J. Classif.*, vol. 17, no. 24, pp. 7-24, 1984.

[23] M. R. Aix-Marseille, "A Comparative Study of Divisive and Agglomerative Hierarchical Clustering Algorithms," *J. Classif.*, vol. 35, pp. 345–366, 2018, Doi: DOI: 10.1007/s00357-018-9259-9.

[24] A. Guénoche, P. Hansen, and B. Jaumard, "Efficient algorithms for divisive hierarchical clustering with the diameter criterion.," *J. Classif.*, vol. 8, no. 1, pp. 5–30, 1991, Doi: doi.org/10.1007/BF02616245.

[25] M. V. Reddy, M. Vivekananda, and R. U. V. N. Satish, "Divisive Hierarchical Clustering with K-means and Agglomerative Hierarchical Clustering," *Int. J. Comput. Sci. Trends Technol.*, vol. 5, no. 5, 2017.

[26] Alok Sharma, Y. López, and T. Tsunoda, "Divisive hierarchical maximum likelihood clustering," *BMC Bioinformatics*, vol. 18, no. 16, pp. 139–147, 2017, Doi: DOI 10.1186/s12859-017-1965-5.

[27] Q. Liu, S. Du, B. Jacobus, V. Wyk, and Sun Yanxia, "Niching particle swarm optimization based on Euclidean distance and hierarchical clustering for multimodal optimization," *Springer Nat. B.V. Nonlinear Dyn.*, vol. 99, no. 3, pp. 2459–2477, 2019, Doi: doi.org/10.1007/s11071-019-05414-7.

[28] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," *Data Min. Knowl. Discov.*, vol. 2, no. 1, pp. 86–97, 2012, Doi: https://doi.org/10.1002/widm.53.

[29] Z. Xu, "Intuitionistic fuzzy hierarchical clustering algorithms," *Journal of Systems Engineering and Electronics*, vol. 20, no. 1. pp. 90–97, 2009.

[30] X. Li, Y. Wang, S. Basu, K. Kumbier, and Bin Yu, "A Debiased MDI Feature Importance Measure for Random Forests," *Neural Inf. Process. Syst. (NeurIPS)*, vol. 32, 2019.

[31] André Altmann, L. Tolosi, O. Sande, and T. Lengauer, "Permutation importance: a corrected feature importance measure," *BIOINFORMATICS*, vol. 26, no. 10, pp. 1340–1347, 2010, Doi: doi:10.1093/bioinformatics/btq134.

[32] C. Arouri, E. Mephu, N. S. Aridhi, C. Roucelle, G. Bonnet-Loosli, and N. Tsopzé, "Towards a constructive multilayer perceptron for regression task using non-parametric clustering. A case study of Photo-Z redshift reconstruction," *ResearchGate*, no. December, 2014.

[33] S. Weisberg, "Yeo-Johnson Power Transformations," *Dep. Appl. Stat. Univ. Minnesota*, no. 2, pp. 1–4, 2001, [Online]. Available: http://stat.umn.edu/arc/yjpower.

[34] Dhiraj K, "Implementing Gradient Boosting in Python," 2019. https://blog.paperspace.com/implementing-gradient-boosting-regression-python.