H & T PUBLISHER

Vol. 3, No. 3 (2022); 75–85

Journal of Earth Sciences and Technology

http://www.htpub.org/Journal-Of-Earth-Sciences-And-Technology/

Role of Supervised Machine Learning for the Prediction of Elastic Logs. A Case Study from Middle Indus Basin, Pakistan

Maha Ali Haider [a]*, Armaghan Faisal Miraj[a], Sher Afgan[a], Rana Faizan Saleem[a], Palwasha Shahzad[a]

[a]Institue of Geology, University of the Punjab, Quaid-e-Azam Campus, 54590 Lahore, Pakistan

Article Information

Article History
Received: 10/05/2022
Accepted: 01/06/2022
Available online: 02/09/2022

Keywords
Supervised Machine Learning
Random Forest
Gradient Booster
Support Vector Machine
Decision Tree

Abstract

In artificial intelligence, machine learning is the branch of the field that can learn from data and recognize patterns in order to make judgments with little or no human interaction. Because of the abundance of easily available data, the petroleum industry is particularly well positioned to benefit from machine learning. This study employed a total of four wells. The purpose of this work is to do the best possible prediction of missing curves which are S-sonic (DTS) with use of four different algorithms of supervised machine learning. The machine learning random forest system was trained on two selected wells in order to forecast the essential DTS curve in missing wells, resulting in an 80% match. Therefore, the use of sophisticated statistical, machine learning, and pattern recognition approaches to solve such challenges has piqued the interest of researchers in the oil and gas industry.

1. Introduction

The study area i.e., Sawan gas field is located, geologically at the fringe of Middle Indus Basin, and geographically situated in Khairpur district of Sindh province, Pakistan. It is well known for one of the propitious gas productions in Pakistan with progressing productivity of around 850 BCF, hence named after it as Sawan gas field. It was tracked down in 1998 in one the famous desert of Sindh i.e., Thar. In this work, four wells of Sawan gas field (Sawan-04, Sawan-05, Sawan-07 and Sawan-08) are used for the study in the prediction of S-sonic b training and testing the data bases of study wells. All these well have LAS files which have the curves of different logs. Among these curves, few curves which are missing in two wells whereas present in other two wells. The sonic log has DT curve for P waves is present in LAS files of all

* Corresponding author
E-mail address: mahaalihaider26@gmail.com

the wells of under study, whereas, the DT curve for S waves is present in two study wells (Sawan-07 and Sawan-08) and absent in other two study wells (Sawan-04 and Sawan-05).

The purpose of this work is to do the best possible prediction of missing curves which are S-sonic (DTS) with use of four different algorithms of supervised machine learning. Two study wells i.e., Sawan-07 and Sawan-08 (both having available data of inputs and outputs) are used as training wells and their data set is used to train the algorithms to drive the relationship between the functions in order to generate the desire result or curve. For this purpose, Sawan-08 was taken as blind well and Sawan-07 as train well to generate DTS curve (as output curve) of Sawan-08 and then correlate it with the measured DTS curve to check the accuracy of the algorithms and then used it for testing wells i.e., Sawan-04 and Sawan-05 for the prediction of elastic logs (S-sonic (DTS)) as it held a great importance in fluid determination. Machine learning is one of the major disciplines of Artificial Intelligence (AI) and emerging over past few decades exponentially due to its wide applications of practical technology with limited data and minimum human intervention.

2. Geology of the Area

Sawan gas field lies in the vicinity of Middle Indus Basin. The middle Indus Basin occupies the eastern border of the Pakistan. The regional geology of the study area is unique as it is bounded by Sargodha High in north, Jacobabad and Mari Kandkot Highs and offshore murray ridges oven fracture plate boundary (Afzal et al., 2009) in south, Indian plate and Kirthar and Sulaiman fold and thrust belts marked the eastern and western boundary of the Middle Indus Basin (Asad & Rahim, 2019). The Sawan field is situated at the south of the Middle Indus Basin (Figure 1). The stratigraphy of the Middle Indus Basin comprises from Triassic age (Wulgai Formation) to Miocene-Pliocene age (Siwaliks) (Siyar et al., 2017). The construction of sawan gas field was mainly governed by following post-rift tectonic events;

- The uplift and erosion during late Cretaceous.
- Thick-skinned wrench faulting, North-West trending.
- The tectonic uplift of Khairpur and Jacobabad highs from late Tertiary till today.

All these above-mentioned tectonic events, especially the tectonic uplift of Khairpur highs (Ashraf et al., 2019) had played a fundamental part in the creation of stratigraphic as well as in structural traps for hydrocarbons, not only in Miano, Kadanwari and Sawan gas fields (Ahmad and Chaudhry, 2002), but also in other oil and gas fields of Middle Indus Basin.

The lithology of the Middle Indus Basin is summarized in column from late Triassic to Miocene-Pliocene in figure 2. On the basis of lithology of the Middle Indus Basin, the Goru Formation of Middle Cretaceous is further subdivided into Upper Goru and Lower Goru (Kadri, 1995). The upper part of Goru Formation has no major reservoir potential as it is mainly composed of shales. However, lower Goru has much more reservoir potential as it mostly composed of sands which are not further glued by shales. The sand part of lower Goru has three main division i.e., upper, middle and basal sands, whereas the upper sands is further divided into sub-zones which are A, B, C and D zones. As zones of upper sand have more zones high potential hydrocarbons (Quadri and Shuaib, 1986). Out of these intervals, C-interval is one of the main importance of sand reservoir.

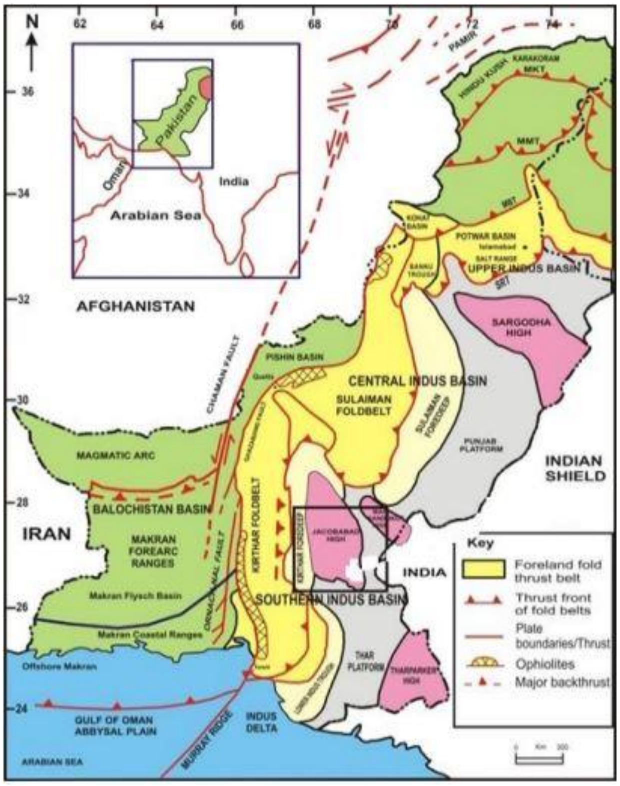

Fig. 1. Map showing the tectonic and sedimentary division and subdivision of geology of Pakistan.

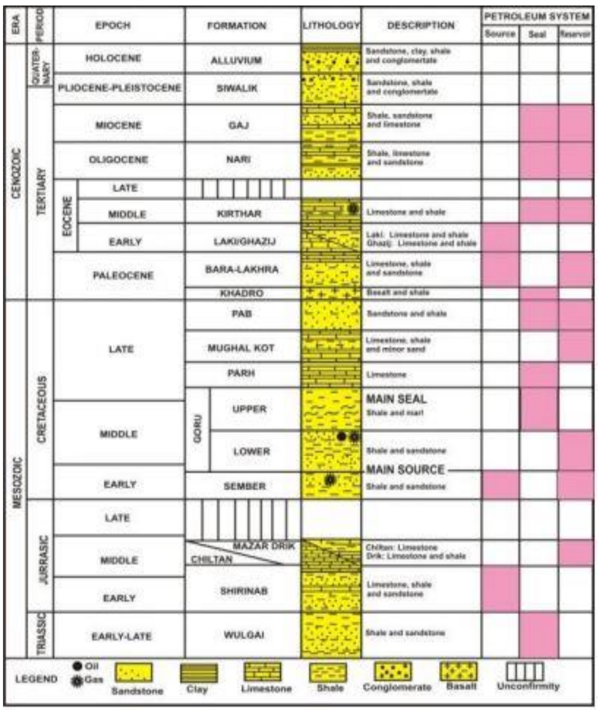

Fig. 2. The summarized stratigraphic column of Middle Indus Basin.

3. Machine learning

Machine learning is setting the new trend of computer based practical applications. It has been progressing exponentially over past few decades and now it engulfed the most of the commercial use of practical technology. It is in fact a branch of artificial intelligence (AI) that trained the data set to perform the relations among data automatically.

It is basically a discipline whose primary objective is related to two interconnected questions i.e., how can the developed computer algorithms can be ameliorate via understanding and practice and the other is to understand which primary information and laws of including both computers and humans are responsible for governing those learning systems. There is a wide range of variety of advanced machine learning algorithms that has been developed in order to hit all the extensive arrays of data and related problem types, which ca appear along several machine learning problems (Jordan and Mitchell, 2015). The machine learning algorithms are basically concerned about the relationship between different functions and tried to learn that relation by training. That relationship is then applied on other functions of testing data to get desired results with minimum human intervention. Therefore, several algorithms of machine learning basically concentrated on the problems of function approximation, where the job is manifested solely in a function, such as providing an input arrangement and labeling the output, then the learning issues can be ameliorating the precision of that function, it can be more enhanced with different practices of a selected training data including pairs of input and output functions.

Following are the main three types of machine learning:

- Supervised Machine Learning
- Unsupervised Machine Learning
- Reinforcement Machine Learning

3.1 Supervised Machine Learning

Supervised Machine Learning is one of the most diverse and important types of machine learning that is greatly used due to its broad variety of multiple functions between input and output data. The dataset or data frame of supervised machine learning are generally categorized or labeled which means that the algorithms that are given can ably determine the attributes with accuracy and also it can carry out predictions on testing data. A feature and a target dataset are used in supervised learning are necessary in order to fit and train the model (Jiang et al.2020).

Supervised learning techniques are frequently capable of handling both regression and classification problems; the method is normally built for one instance and changed for the other (Caté et al.2017). The supervised machine learning has 4 major types of algorithms, which are designed specifically for certain task, which are as follows:

- Random Forest (RF)
- Support Vector Machine (SVM)
- Decision Trees (DTR)
- Gradient Booster (GB)

4. Materials and Methods

In this work, four different algorithms of supervised machine learning are used to train the data set of available inputs and outputs of training data and then applied on the testing data of missing outputs for the prediction of outputs, that is based on the resulting accuracy of outputs of training data. For the QC of input data which are in LAS format have been done by the generation of heatmaps and boxplots which are helpful in checking the faulty data at one glance.

5. Results and Discussions

As there are two types of data i.e., training data and testing data. The Sawan-07 and Sawan-08 are two wells that are used in training data and has all the 5 available input and output curves, used for the prediction of S-sonic. The GR, RHOB, NPHI and DT are used as input curves for S sonic prediction in Sawan-04 and Sawan-05 wells.

5.1 Input dataset of machine learning for Sawan gas field

The training datasets are those data set which are used to train machine learning models. The algorithms of machine learning performed desired task and do predictions with the help of training data sets. In this study, the GR, RHOB, NPHI and DT are the input curves of Sawan-04 and Sawan-05 wells (Figure 3) for s-sonic prediction. Whereas, the GR, DT, NPHI, RHOB and DTS are the available curves of Sawan-07 and Sawan-08 wells (Figure 4), these data sets are used for training. In short, in this work the data sets of Sawan-07 and Sawan-08 are training data and the data set of Sawan-04 and Sawan-05 are testing data.

5.2 Heatmap and boxplots of data

In order to QC the data at first glance, the data is used to generate heatmaps and boxplots to easily foresee the graphical display of the data in terms of colors. This is useful in understanding the complex data through visual aid and to check the misread and faulty readings of the data easily. The heatmaps and boxplots of Sawan-04 (Figure 5), Sawan-05 (Figure 6), Sawan-07 (Figure 7) and Sawn-08 (Figure 8) wells are generated by using the input curves of GR, RHOB, NPHI, DT and DTS both from training and testing data. In these boxplots there are some outliers but due geological features we cannot remove the outliers. Using these plots we analyze the minimum, maximum and standard deviation of the datasets.

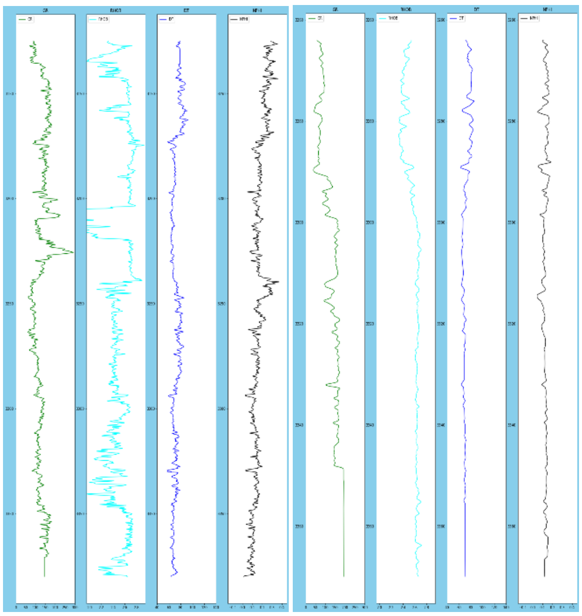

Fig. 3. Input data sets of machine learning for Sawan-03 and Sawan-04 wells of Sawan gas field and shown the curves of GR, RHOB, DT and NPHI in this order.

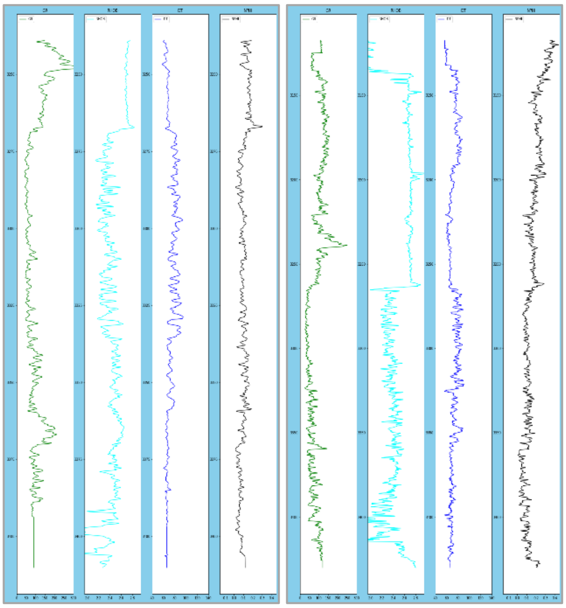

Fig. 4. Input data sets of machine learning for Sawan-07 and Sawan-08 wells and shown the curves of GR, RHOB, DT and NPHI in this order, that is used in training for the prediction of DTS.

a) b)

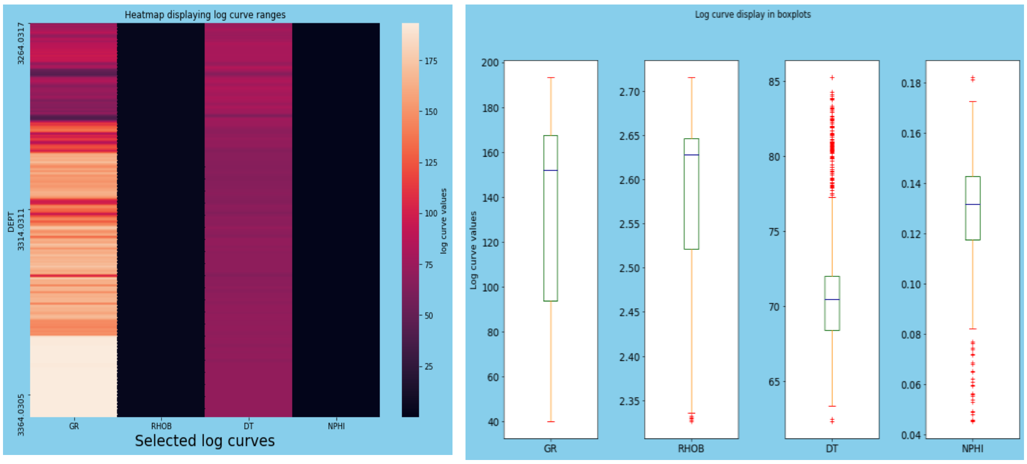

Fig. 5. a) Heatmap of Sawan-04 shows values of GR, RHOB, DT and NPHI using colors b) Boxplot of Sawan-04 shows different quantiles i.e., minimum, maximum and standard deviation of GR, RHOB, DT and NPHI curves.

a) b)

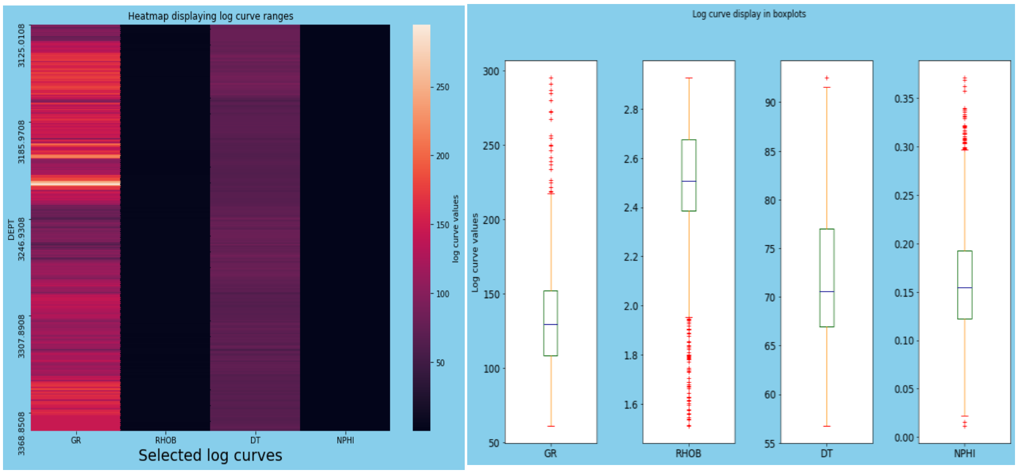

Fig. 6. a) Heatmap of Sawan-05 shows values of GR, RHOB, DT and NPHI using colors b) Boxplot of Sawan-05 shows different quantiles i.e., minimum, maximum and standard deviation of GR, RHOB, DT and NPHI curves.

a) b)

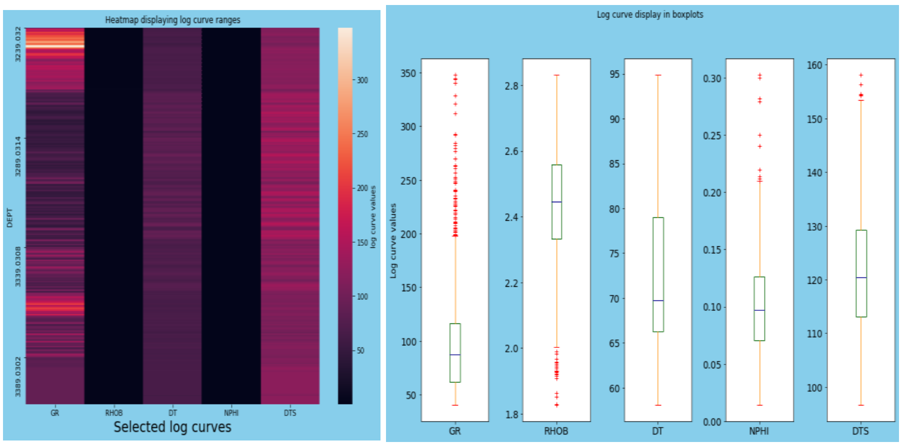

Fig. 7. a) Heatmap of Sawan-08 shows values of GR, RHOB, DT, NPHI and DTS using colors b) Boxplot of Sawan-08 shows different quantiles i.e., minimum, maximum and standard deviation of GR, RHOB, DT, NPHI and DTS curves.

a) b)

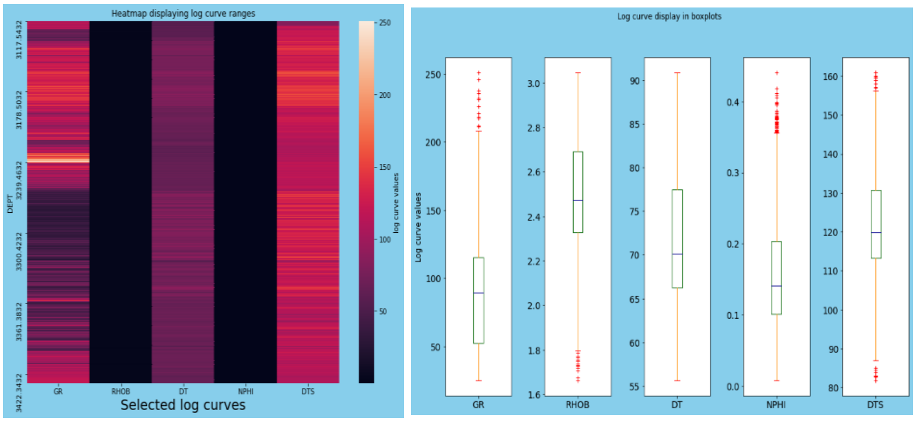

Fig. 8. a) Heatmap of Sawan-07 shows values of GR, RHOB, DT, NPHI and DTS using colors b) Boxplot of Sawan-07 shows different quantiles i.e., minimum, maximum and standard deviation of GR, RHOB, DT, NPHI and DTS curves.

5.3 Prediction of S-sonic in Sawan-08 well

In order to select the best predicted s-sonic for previous mentioned wells, the training of Sawan-07 and Sawan-08 wells was done by taking the Sawan-08 well as a blind well and Sawan-07 well as training well to generate the predicted s-sonic for Sawan-08, by all four algorithms of supervised machine learning and then correlated it with already measured s-sonic (DTS). The correlation scores for all algorithms are given in the table 1. The correlation match of S-sonic (DTS) curves of both predicted and measured Sawan-08 well (Figure 9) also shows the visual correlation of the curves by all four algorithms.

Table 1. R2 Score (%) for prediction S-sonic by correlating it with measured S- sonic at Sawan-08, for Random Forest (RF), Decision tree (DTR), Support vector regression (SVR) and gradient boost (GB).

Algorithm	Random Forest (RF)	Decision Tree (DTR)	Support Vector Regression (SVR)	Gradient Booster (GB)
R2 Score (%)	82	73	81	81

5.4 Prediction of S-sonic in Sawan-04 and Sawan-05 wells

Four algorithms of supervised machine learning ware used, by taking the input curves i.e., GR, NPHI, RHOB, DT and DTS of both Sawan-07 and Sawan-08 wells as training wells and run the algorithms turn by turn on testing wells of Sawan-04 and Sawan-05. The data sets of Sawan-07 and Sawan-08 both are collectively used as trained data for Sawan-04 and Sawan-05 individually, in order to predict the S-sonic of both wells separately. The predicted S-sonic of Sawan-04 is shown in figure 10a and Sawan-05 in figure 10b, respectively.

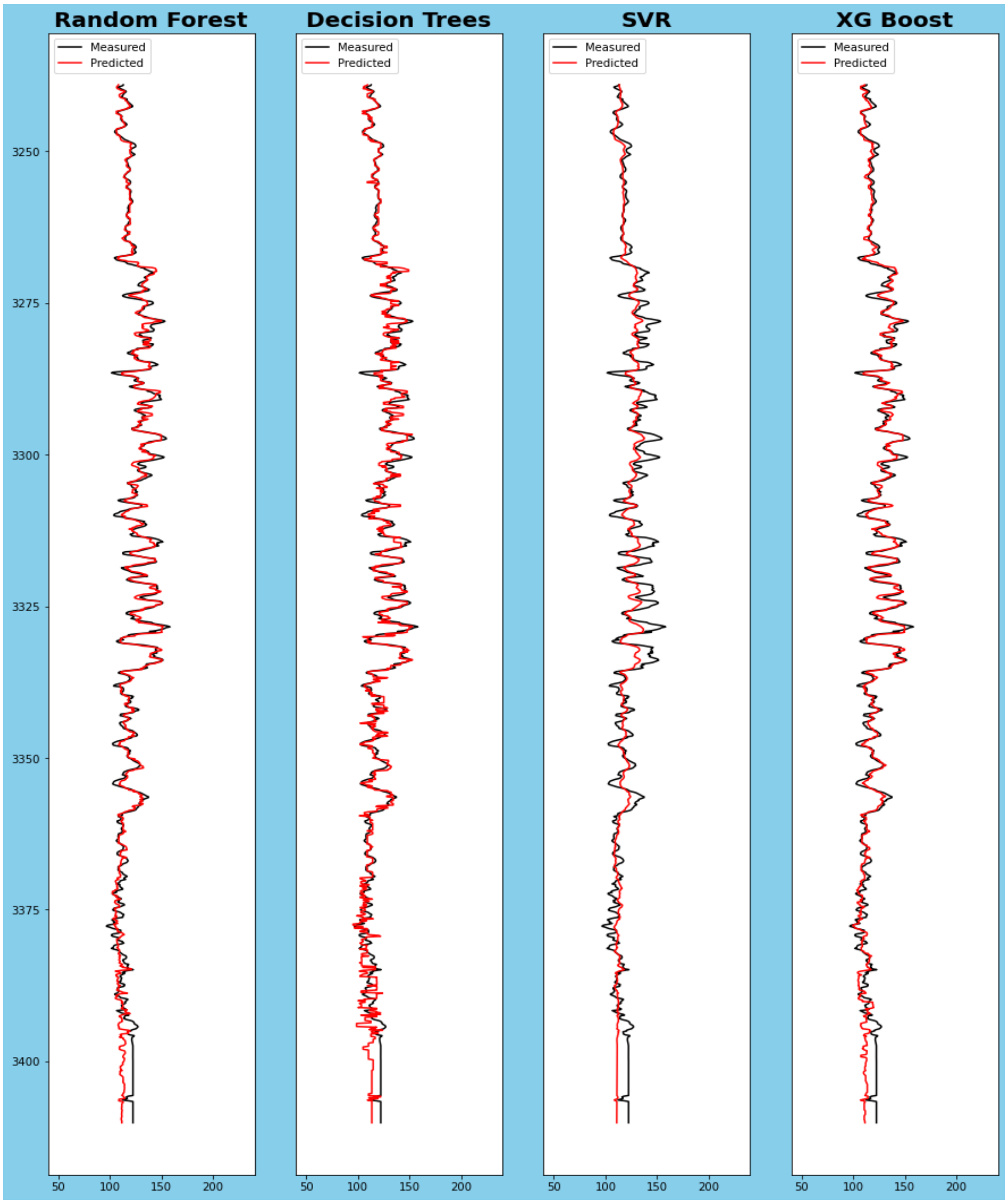

Fig. 9. Prediction of S-wave at Sawan-08 using Random Forest, Decision Tree, Support Vector regression and gradient boost and correlated it with measured sonic.

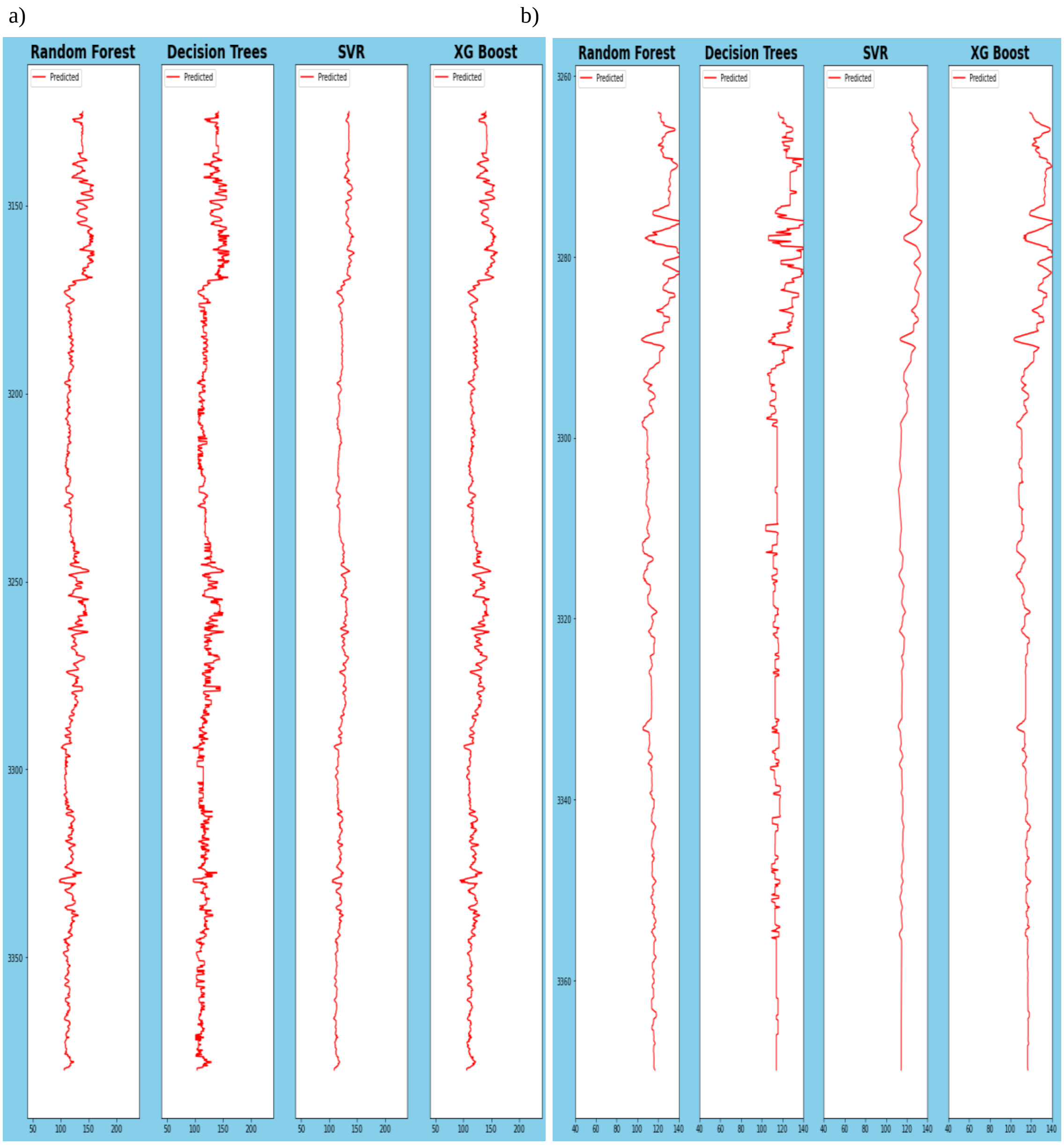

Fig. 10. a) Prediction of s-sonic at Sawan-04 using Random Forest, Decision Tree, Support Vector Regression and gradient boosting, b) Prediction of s-sonic at Sawan-05 using Random Forest, Decision Tree, Support Vector Regression and gradient boosting.

6. Conclusions

The given work is all about the prediction of missing elastic log with the use of supervised machine learning in certain old wells of Sawan Gas Field in Middle Indus Basin. The sonic log for p-waves is present in all four study well whereas the sonic log of s-waves known as s-sonic (DTS) is present in two study wells i.e., Sawan-07 and Sawan-08 and absent in remaining two study wells i.e., Sawan-04 and Sawan-05. As s-waves plays an important role in fluid identification hence the importance of S-sonic in the interpretation of porosity of the formation and to give required information in order to support and calibrate the seismic data as well. The S-sonic measures the transit time (Δt) which is microsecond per feet. The S-sonic in old wells such as in Sawan-04 and Sawam-05 was not acquired and cannot be logged in, therefore in order to overcome the problem, the four algorithms of supervised machine learning were used on training wells i.e., Sawan-07 and Sawan-08 (having S-sonic (DTS)). Firstly, Sawan-08 was considered as bind well, and tested against trained well of Sawan-07, for the prediction of DTS and then correlated it with measured DTS to check out the accuracy of one out of four algorithms of supervised machine learning (Random Forest, Support Vector Regression, Decision Tree, Gradient Booster) that were used. The results of correlations of all algorithms were shown both in the form of R2 score (%) and in curves to single the best predicted algorithm. Among them Random Forest (RF) shows the beast prediction. Then, moved on to the prediction of S-sonic in testing wells (Sawan-04 and Sawan-05) by using the algorithms. As random forest was the best match in Sawan-08, therefore it also showed the better prediction result for s-sonic in both testing wells. Hence, with limited data supply, machine learning is the most diverse and applicable to wide arrange of data sets is able to provides the best results by training and testing data, also in limited supply of data.

References

Asad, M., & Rahim, H. U. (2019). Porosity Distribution and Differentiation of Different Types of Fluids in Reservoir of Sawan Gas Field, Lower Indus Basin, Pakistan. *Pakistan Journal of Geology*, 3(1), 28-37.

Afzal, J., Kuffner, T., Rahman, A., & Ibrahim, M. (2009). Seismic and well-log based sequence stratigraphy of the early Cretaceous, Lower Goru “C” sand of the Sawan gas field, middle Indus Platform, Pakistan. In Proceedings, Society of Petroleum Engineers (SPE)/Pakistan Association of Petroleum Geoscientists (PAPG) Annual Technical Conference, Islamabad, Pakistan.

Quadri, V. U. N., & Shuaib, S. M. (1986). Hydrocarbon prospects of southern Indus basin, Pakistan. *AAPG bulletin*, 70(6), 730-747.

Siyar, S. M., Waqas, M., Mehmood, S., Jan, A., Awais, M., & Islam, F. (2017). Petrophysical characteristics of lower Goru formation (Cretaceous) in Sawan gas field, central Indus basin, Pakistan. *Journal of Biodiversity and Environmental Sciences (JBES)*, 10(5), 260-266.

Ashraf, U., Zhu, P., Yasin, Q., Anees, A., Imraz, M., Mangi, H. N., & Shakeel, S. (2019). Classification of reservoir facies using well log and 3D seismic attributes for prospect evaluation and field development: A case study of Sawan gas field, Pakistan. *Journal of Petroleum Science and Engineering*, 175, 338-351.

Kadri, I. B. (1995). Petroleum geology of Pakistan. Pakistan Petroleum Limited.

Ahmad, N., & Chaudhry, S. (2002). Kadanwari gas field, Pakistan: a disappointment turns into an attractive development opportunity. *Petroleum Geoscience*, 8(4), 307-316.

Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020). Supervised machine learning: a brief primer. *Behavior Therapy*, 51(5), 675-687.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.

Caté, A., Perozzi, L., Gloaguen, E., & Blouin, M. (2017). Machine learning as a tool for geologists. *The Leading Edge*, 36(3), 215-219.