**Estimation of Missing Petrophysical Data through a Machine Learning Bagging Approach**

**AIM**

* Process and reproduce a full and complete data for petrophysical interpretation

**OBJECTIVES**

* Estimate missing logs in the incomplete wells
* Estimate missing reservoir sections
* Evaluate and validate methods used
* Check for predicted logs correlation

**ABSTRACT**

Proper evaluation of petrophysical parameters is very crucial in the formation evaluation of a reservoir. Calculated with the help of well logs, these calculated parameters and values are subjected to a number of uncertainties arising from the log acquisition process, logging devices and calibrations, heterogeneity nature of the subsurface. Missing values and incomplete well data as well aid the complexity associated with reducing these uncertainties reducing the quality of interpretation done to evaluate the reservoirs. The aim of the study is to present a machine learning approach to solve for missing values using bootstrap aggregation (bagging).

Bootstrap aggregation is a machine learning ensemble technique used to average the predictions of multiple tree-based learners. This helps to reduce variance in prediction of the test data without increasing the bias, thus preventing overfitting of the training data. Two bagging algorithms were used (Random Forest Regressor and Extra Trees Regressor) and a single Decision Tree Regressor. Six wells from a field in the North Sea were used. The petrophysical parameters of the wells are the Delayed Time (DT), Gamma Ray (GR), Porosity (NPHI), Bulk Density (RHOB) and Resistivity (RT). Four of the wells had the complete parameters and these four were used for the training and validation process. The other two wells served as the test data used for final predictions for the missing DT logs. Missing sections on the NPHI and RHOB correlating to the reservoir region on a well log were also estimated using the same method. Hyperparameter tuning of the algorithms used were done to get the optimal hyperparameters for the training process. Cross validation of the models was also done for performance evaluation.

The machine learning models were used in estimating the delayed time log on the validation datasets. The bagging models yielded an R2 score in the range of 0.82 to 0.93 and RMSE in the range of 4.09 to 5.72 for delayed time response of the three validation wells used. This showed significance difference when compared to the single decision tree used. However, from the R2 scores and RMSE scores obtained, the Extra Trees Regressor had a better performance than the Random Forest Regressor. The Delayed Time response for the missing wells 11B and 15D were finally predicted using the two bagging algorithms. The missing NPHI and RHOB sections in Well F12 were also estimated using the Extra Trees Regressor. These results and study demonstrate the reliability and accuracy of bagging as a machine learning technique for solving missing well information problems as opposed to the single decision tree technique.

**Summary**

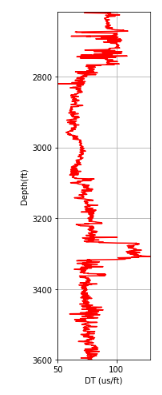
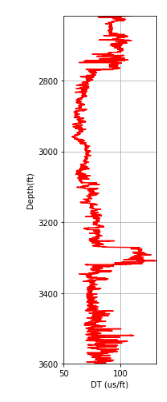
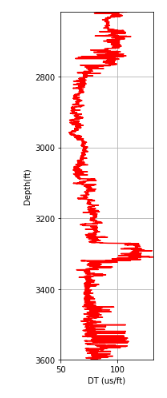
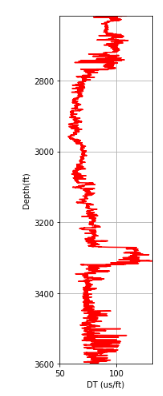
* 6 wells were used;
  + 4 for train and validation dataset
  + 2 for the actual prediction (test dataset)
* 3 models were used
  + 2 Bagging Algorithms used (Random Forest Regressor and Extra Trees Regressor)
  + 1 Decision Tree Regressor
* Evaluation metrics
  + RMSE
  + R2 scores
* Choosing the best model criteria
  + Evaluation metrics
  + Log/Signature similarity of training dataset with test dataset
* Missing reservoir section
  + NPHI (Porosity)
  + RHOB (Density)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Well Names** | **Evaluation Metric** | **Random Forest Regressor (RFR)** | **Extra Trees Regressor (EXT)** | **Bagged Average**  **(RF and EXT)** | **Decision Tree Regressor (DTR)** |
| **F1** | R2 Score | 0.82 | 0.83 | 0.83 | 81.77 |
| **F1** | RMSE | 5.72 | 5.49 | 5.54 | 5.86 |
| **11A** | R2 Score | 0.89 | 0.90 | 0.90 | 88.77 |
| **11A** | RMSE | 4.74 | 4.57 | 4.62 | 4.78 |
| **T2** | R2 Score | 0.91 | 0.93 | 0.92 | 87.00 |
| **T2** | RMSE | 4.57 | 4.09 | 4.22 | 5.63 |

**VALIDATION**

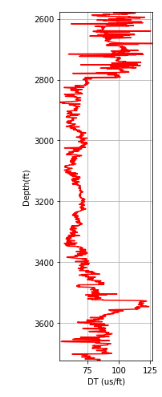
**WELL F1**

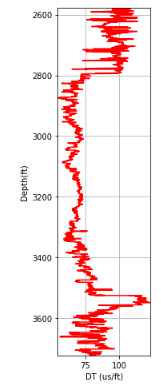
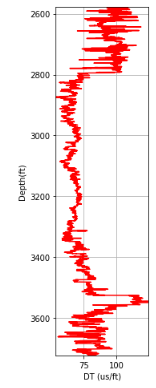
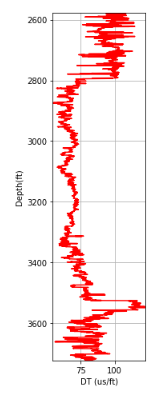
**ACTUAL RESPONSE EXT RFR AVG**

****

**WELL 11A**

**ACTUAL RESPONSE EXT RFR AVG**





**WELL T2**

**ACTUAL RESPONSE EXT RFR AVG**

