PIONEERING THE DRILL:
MACHINE LEARNING FOR PREDICTING RATE OF PENETRATION IN DRILLING OPERATIONS

Akash Borigi
University of Houston
aborigi@uh.edu

Rahul Krishna Gunneri
University of Houston
rgunneri@uh.edu

Dvijesh Shastri PhD
University of Houston-Downtown
shastrid@uhd.edu

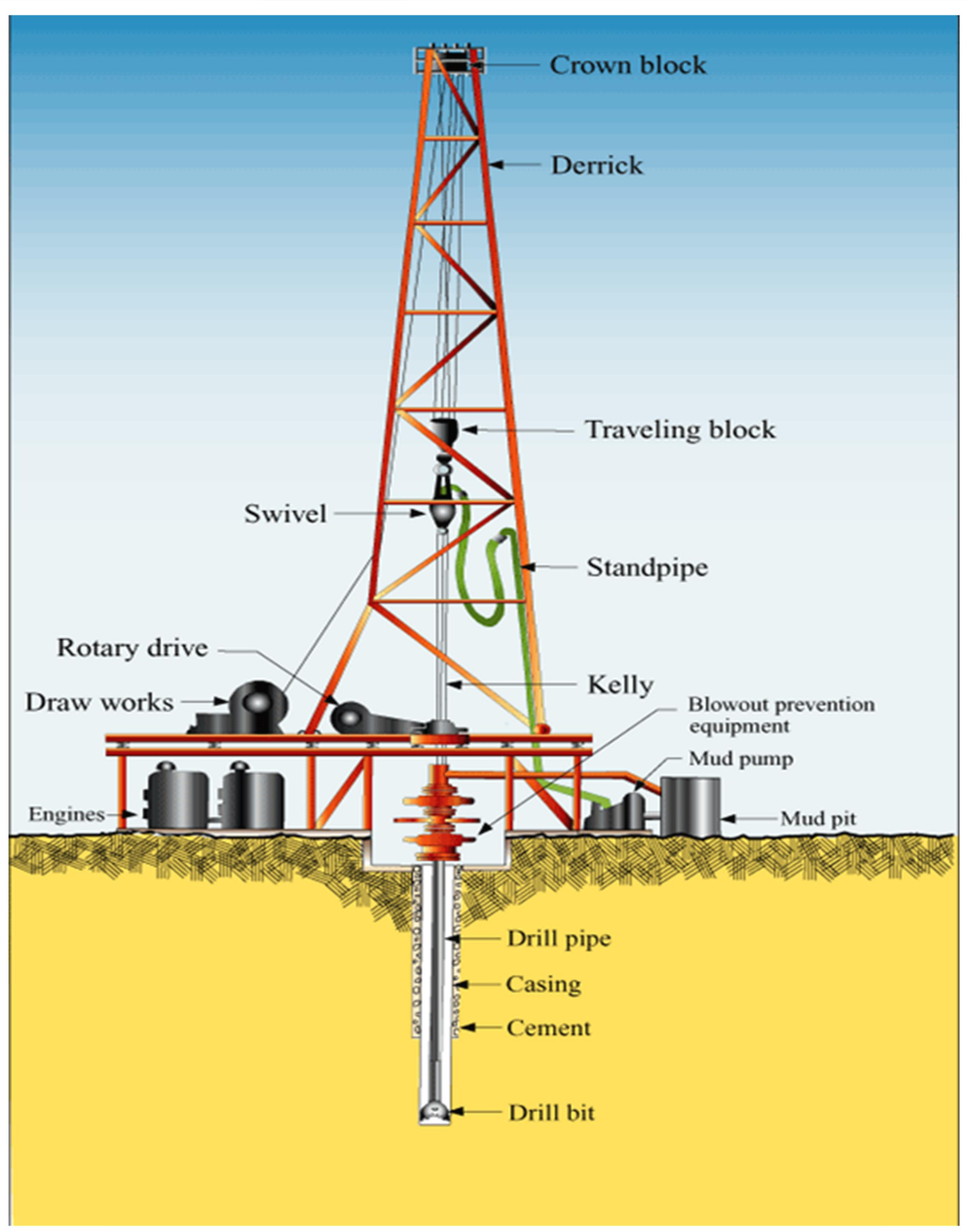
Kalyan Venugopal PhD
University of Wyoming
kalyan.venugopal@uwyo.edu

SCAN ME

ABSTRACT

In the oil and gas industry, the rate of penetration (ROP) is a critical parameter that directly impacts drilling efficiency and cost-effectiveness. Traditional methods for predicting ROP rely heavily on human expertise and historical data, often leading to suboptimal outcomes due to the complexity and variability of drilling conditions. Leveraging the power of machine learning, this study proposes a novel approach to predict ROP with higher accuracy and reliability. By integrating real-time sensor data, geological information, and drilling parameters, advanced machine learning algorithms such as simple computational and supervised machine learning methods are employed to model the intricate relationships between drilling variables and ROP. Therefore, it revolutionizes drilling operations by optimizing efficiency, reducing costs, and enhancing overall productivity in the oil and gas sector.

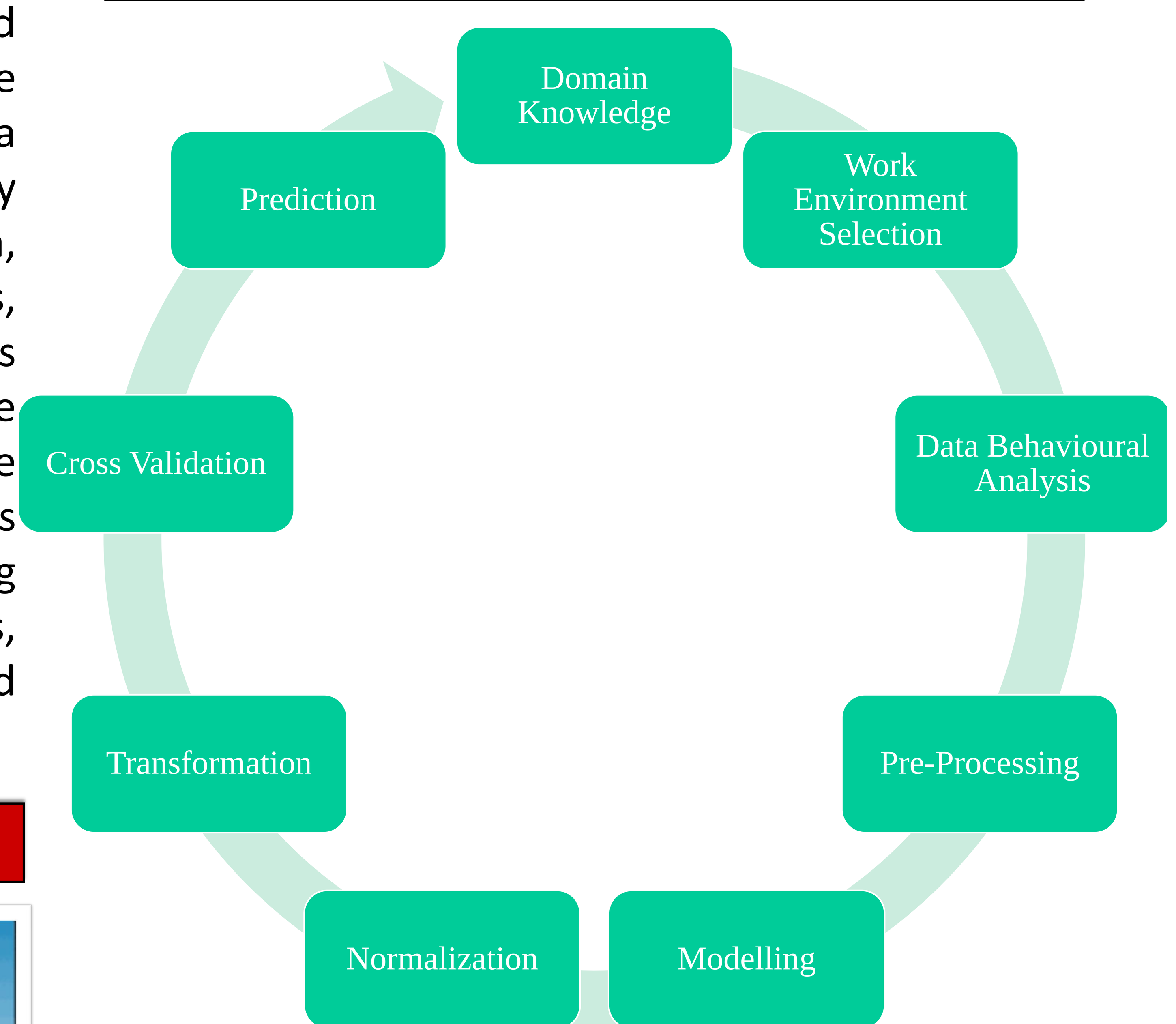
BACKGROUND



OBJECTIVE

"Developing a machine learning model to predict drilling rate of penetration in oil and gas exploration for enhanced operational efficiency."

METHODS



Modeling Techniques:

1. Linear Regression
2. Decision Tree
3. Random Forest
4. K Nearest Neighbor

Normalization:

1. Min-Max Scaling
2. Z-Score Standard Scaling

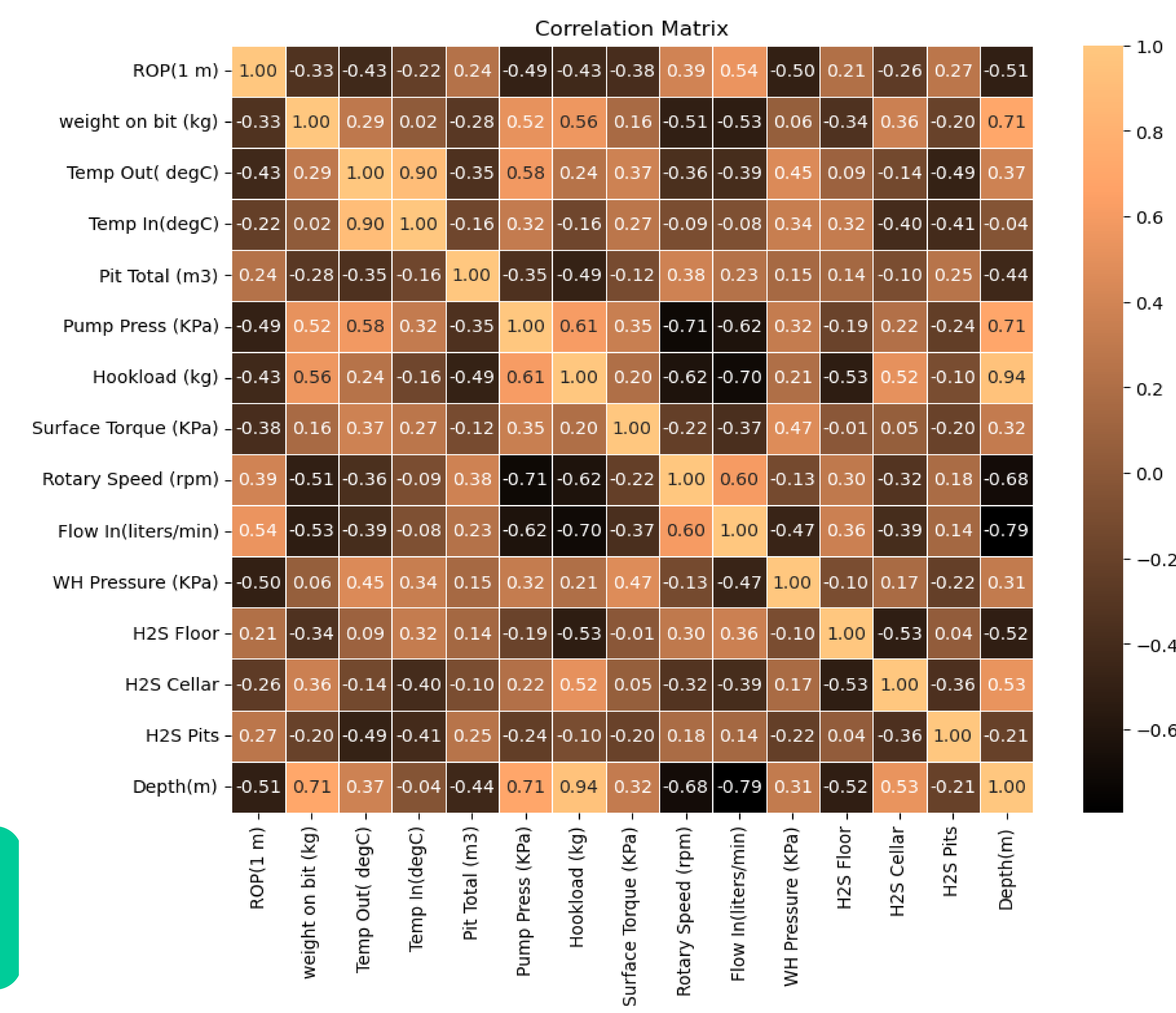
Transformation:

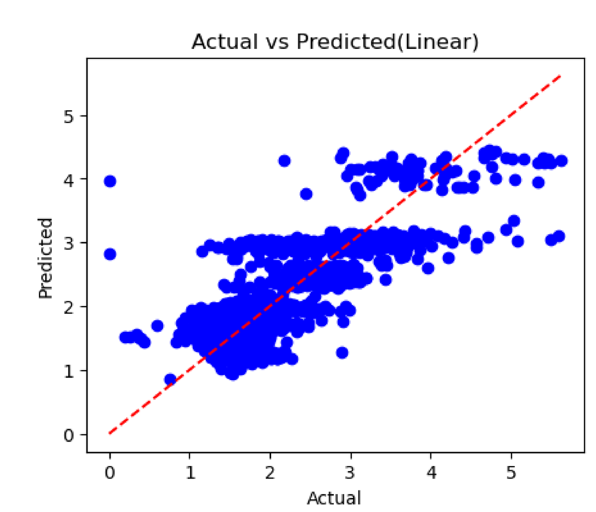
1. Log
2. Square Root
3. Cube Root

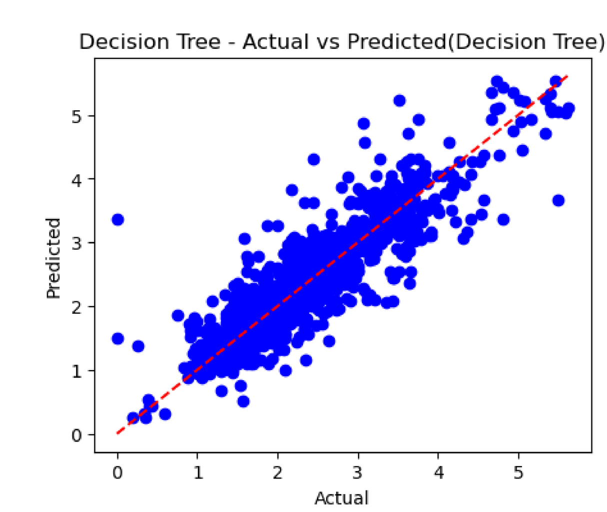
Cross Validation:

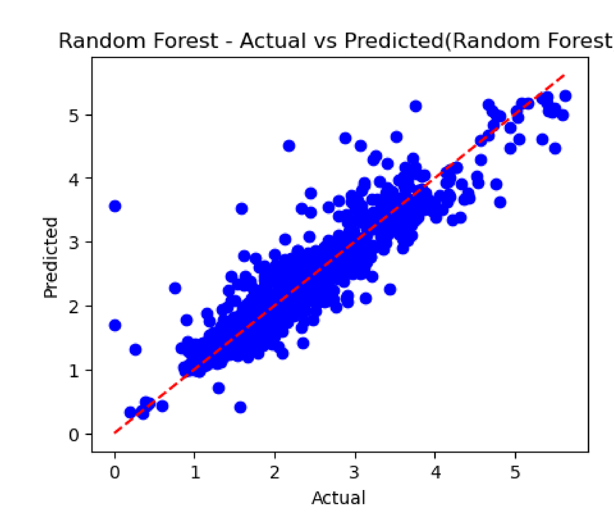
Considered up to 10 splits for cross-validating the training set.

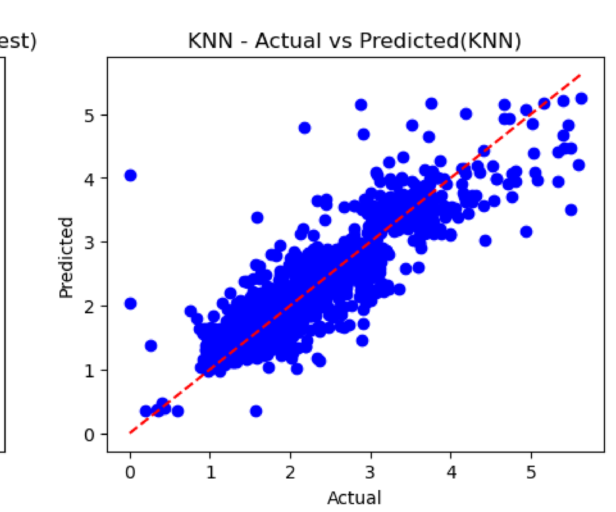
EVALUATION

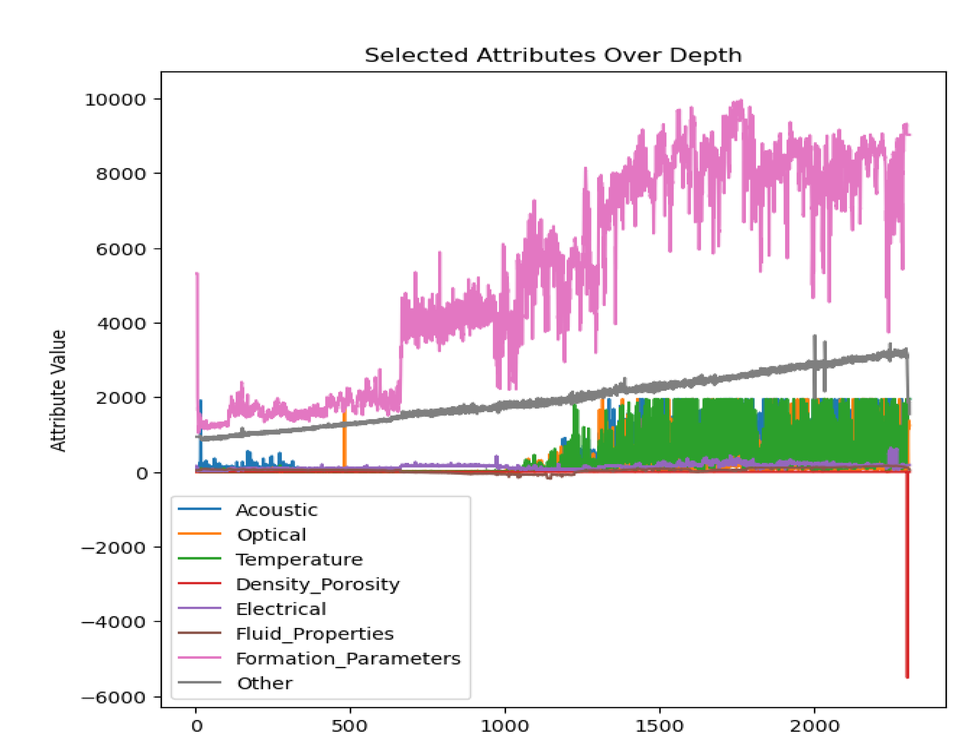






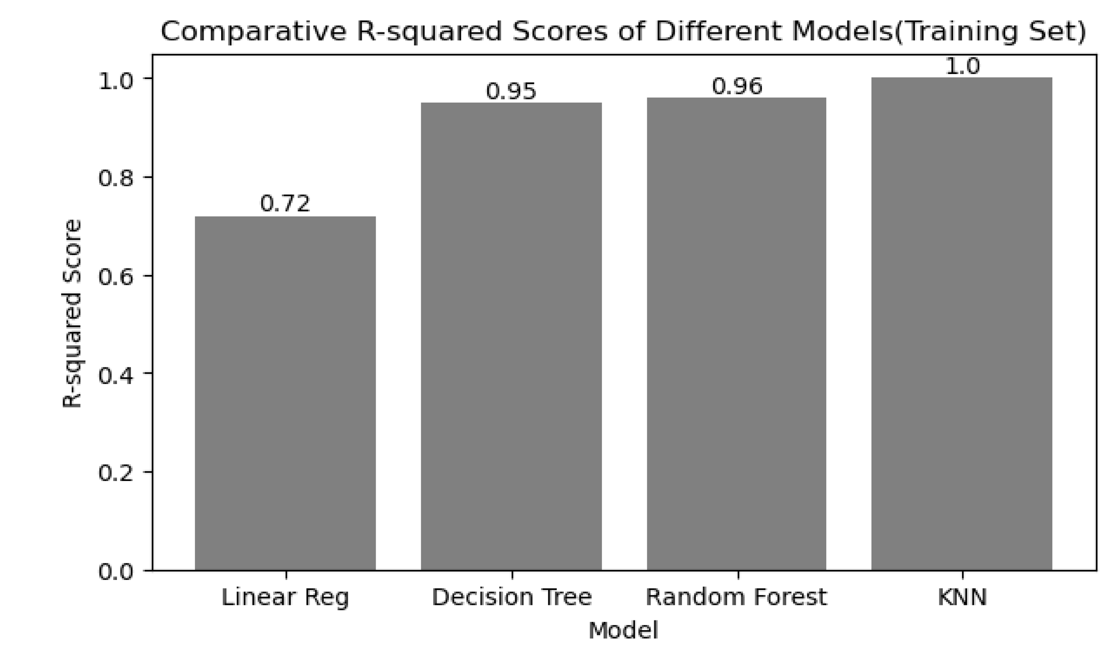


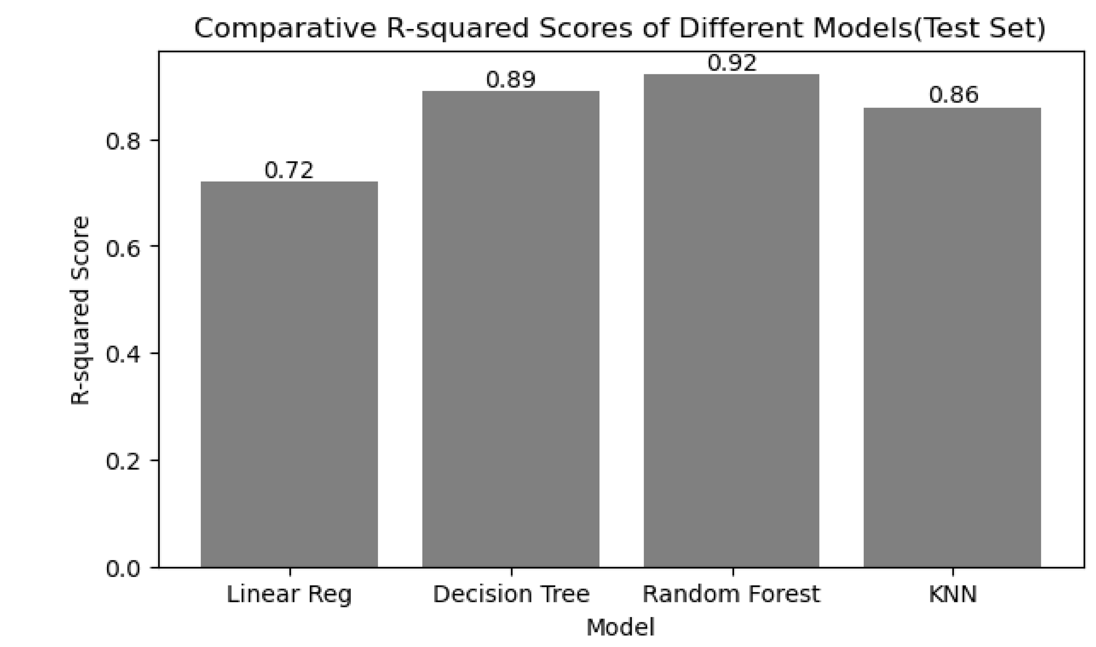




RESULTS

Model	Training RMSE	Test RMSE
Linear Regression	0.45	0.47
Decision Tree	0.19	0.29
Random Forest	0.17	0.25
KNN	0.0	0.33





Integrated the ROP dataset with the geophysical well dataset for modeling and observed the same R-squared value using the Random Forest model due to its high performance in the test set. Included PCA (n-component=1) for data reduction technique while integration for each category.

CONCLUSIONS

"Random Forest outperformed other models with an RMSE of 0.25 and R-squared of 0.92. Future work includes deep learning methods and feature engineering for enhanced accuracy."

Note: For Datasets and References please view the above provided QR code.

ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation under Grants numbers NSF IIS-2123247. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.