Rate of Penetration Prediction Using Machine Learning Models

MJ D Asuncion
University of Houston-Downtown
mjasuncion@acm.org

Dalton Carter
University of Houston - Victoria
daltocarter@outlook.com

Gopika Kizhuvettil
University of Houston-Clear Lake
gopika751@gmail.com

Taven Tran
University of Houston
taventran@gmail.com

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

Kalyan Venugupal PhD
Industry Professional
kalyan100673@gmail.com

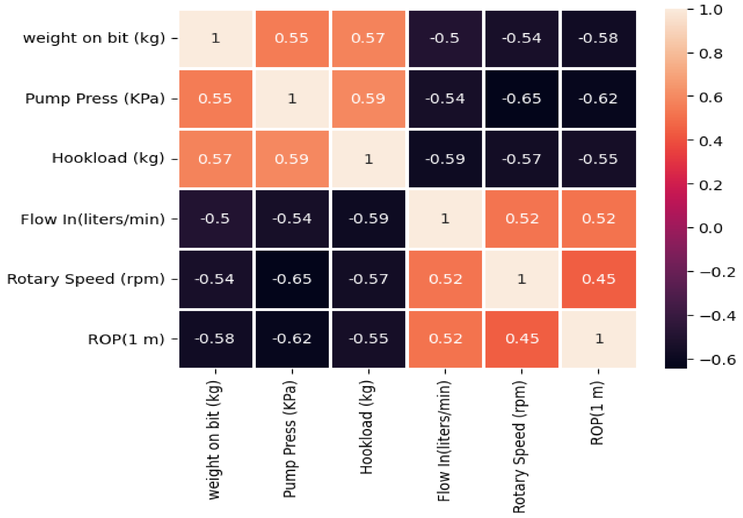
Abstract

The success of drilling a well for geothermal, oil, and gas purposes are quantified by how quickly the desired depth is reached. The progress of a drilling operation is determined by the drill bits rate of penetration (ROP). In our research we use machine learning models to make the drilling process more efficient by predicting the ROP considering the weight on bit (WOB), revolutions per minute (RPM), surface torque, and flow rate. Using data from [1], we cleaned, integrated, and performed exploratory data analysis to train regression models that we evaluated and improved. We concluded that the model best applicable is Random Forest because of its efficiency in training time, memory consumption, and simplicity.

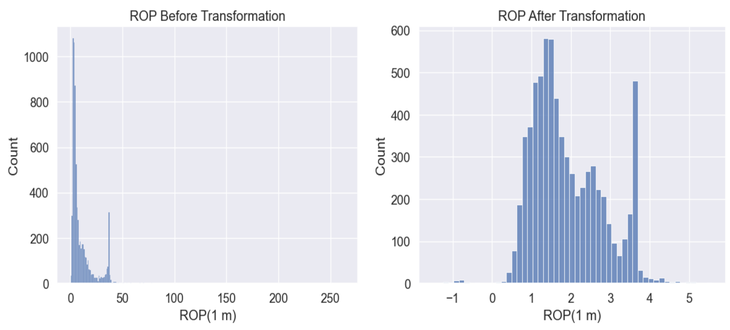
Motivation

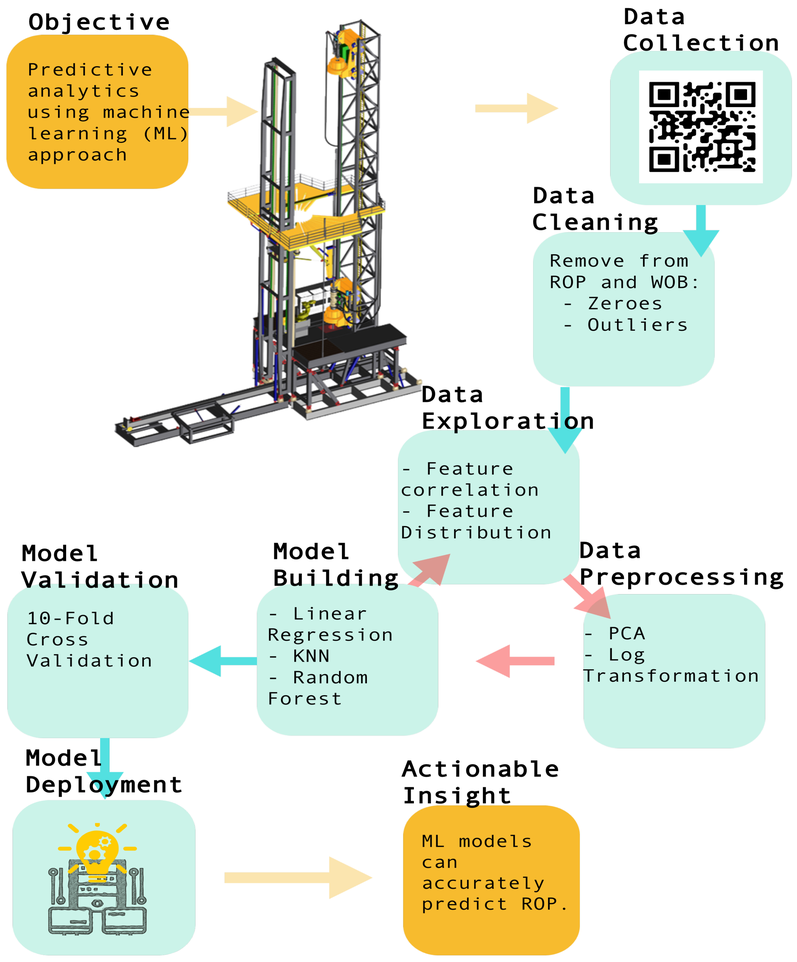
- Optimize ROP prediction
  - Train machine learning models to predict ROP in drilling operations
- Decide the best model to deploy
  - Highest score values
  - Least computational time

Correlation Analysis



Log Transformation

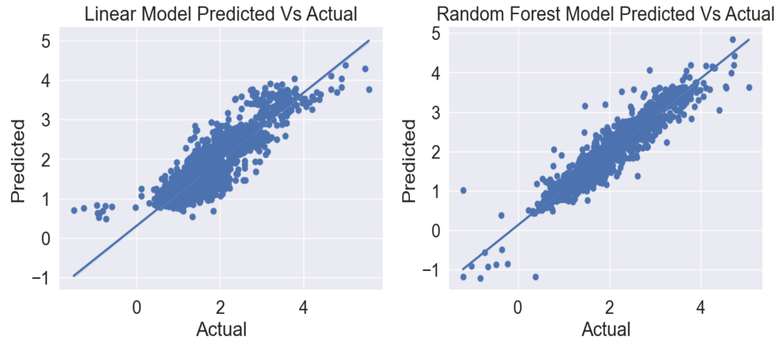




Results

After evaluating several machine learning models on our dataset, we have identified *RandomForest* as the best model for our dataset. This model achieved the highest R2 score of 0.9358 and the lowest RMSE of 0.2320. Random Forest is an ensemble learning method that uses multiple decision trees to make predictions, which results in a more accurate and stable model compared to other models. In our case, we used 200 trees with the Gini impurity criterion. We recommend using this model for future predictions on our dataset.

Model Evaluation



Results Summary

Model	Train Time	Predict Time	Memory	R2	RMSE	Parameters
Linear Regression	0.05s	0.00s	0.200	0.8558	0.3503	
PCA	0.27s	0.00s	0.027	0.8324	0.3732	Features reduced from 73 to 42
KNN	0.00s	0.97s	00.383	0.9087	0.2694	K=4, metric=minkowski, weights=distance
Random Forest	382.08s	0.13s	74.719	0.9358	0.2320	trees=200, gini
Neural Networks	10.64s	0.13s	39.700	0.9020	0.2826	Total Params=19,969, Epochs=101, 4 dense layers
SVM	149.84s	0.66s	90.250	0.9190	0.2607	C=100, gamma=0.1

Processor: Intel(R) Core(TM) i5-9600K CPU @ 3.70GHz 3.70 GHz
Installed RAM: 16.0 GB
System type: 64-bit operating system, x64-based processor

Conclusion & Future Work

Random forest produced results with the least amount of errors out of every model utilized. One feature to be incorporated for future work would be depth when dealing with data from more than one oil well. The model could then be better utilized for predictions on larger data sets from multiple oil well sources.

References

[1] McLennan John Moore Joe Simmons Stuart Wannamaker Phil Allis Rick Podgorney, Robert and Clay. Jones. Utah forge: Well data for student competition.

Acknowledgements

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.