Chevron Rice Datathon Track:

Predicting Rate of Penetration (ROP)

McWiess

Emma Dunn, Steven Feng, William Lu, Shreya Nidadavolu

Executive Summary

Goal: Develop a model to predict rate of penetration based on pre-determined and controllable variables.

Approach: Transform and clean data to combat skewness and multicollinearity.

Results: Predictive model that combines both quantitative and qualitative variables to estimate rate of penetration.

Conclusion: Our model is accurate and representative of the data, but other factors must also be taken into account when determining ROP.

Our approach to the data

Guide our general picture of the dataset with basic linear regression

- Starting with relatively clean data, so can provide a good idea of what we're working with.
- Determined data is skewed right.

Filter out variables based on importance to answering the question

- ROP does not depend on rock depth (max - min) of site, and max and min are correlated; chose to use max b/c worst case analysis.
- Segment ID does not provide information on environmental or other factors that can affect ROP; simply identifies the location.

Transformation and Outliers

- Dealing with non-numeric data.

Considerations of the model

log(ROP)

Overfitting

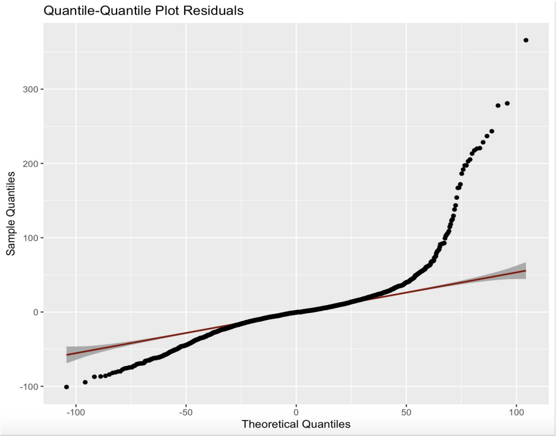
Levels of significance

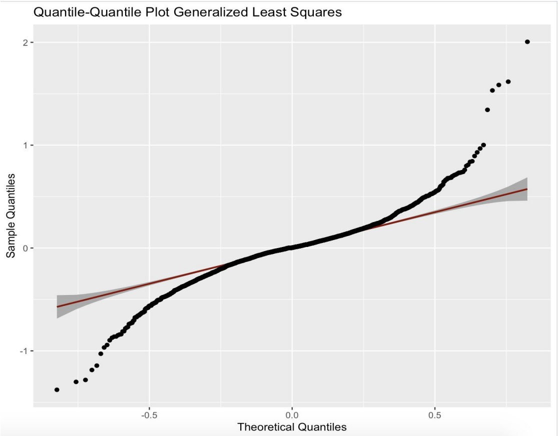
Insignificant independent variables

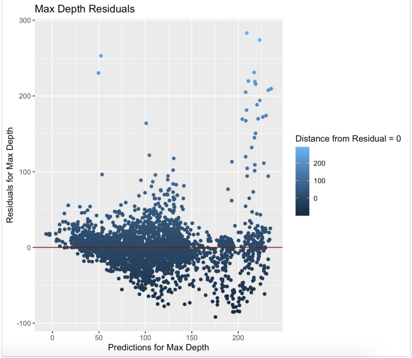
Heteroscedasticity

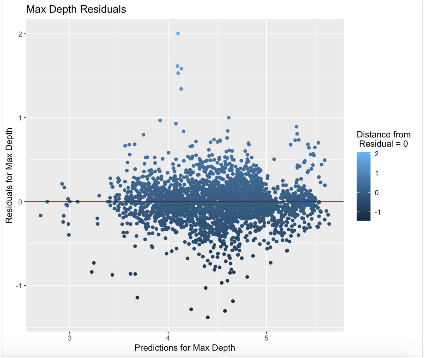
Multicollinearity

Model development









Final model

$$\text{Rate Of Penetration} = X\widehat{\beta}_{WLS}$$

where X is design matrix of predictor variables:

wellbore_chev_no_id , formation_id , drillbit_size , max_depth , surface_weight_on_bit , surface_rpm

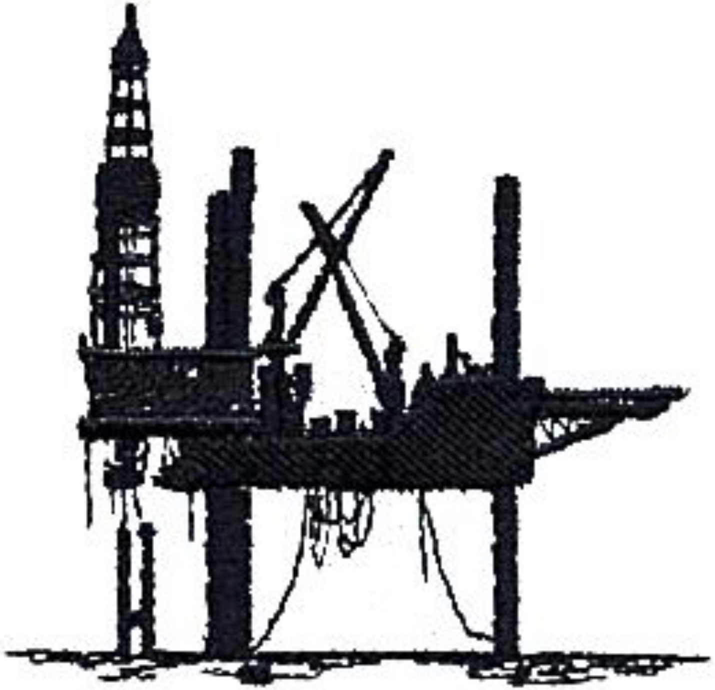
and $\widehat{\beta}_{WLS}$ is the GLS estimate of coefficients.

Model against scoring.csv

Output of model with scoring.csv data

Predictive analysis (RMSE)

Visualization



Possible externalities to consider

- Costs
  - Can use model to determine factors for a slower rate of penetration for shallower depths of rock formations.
    - Not as necessary for a fast rate of penetration because relatively less rock to drill through.
- Social impact
  - Costs of positive marketing for new drilling sites?
- Environmental impact
  - Does increased rate of penetration result in increased risk of errors?