Rod Pump Failure Analysis: Final Presentation

Preston Hart, Rohan Sura, John Mason, Ali Modak, and Tarini Thiagarajan
NSC 325, Fall 2020

TEXAS

ConocoPhillips

What is a Rod Pump?

Sucker Rod Pumps are a type of artificial lift system used to drill for oil.

There is a large, international market for Sucker Rod Pumps consisting of several manufacturers and drilling companies.

The Rod Pump Challenge:

The lost revenue from a single broken rod pump could be as much as $5700 per day.

Some Sucker Rod Pumps are in remote or hard to access locations leaving fewer opportunities for maintenance and repairs.

Our Solution:

We aimed to **identify** and **understand** factors which contribute to Rod Pump failure, and create a product which helps **maximize pump run time** and **minimize cost**.

Throughout this process, we have **developed a framework for Rod Pump telemetry analysis** which can be applied to a pumps in additional regions.

What have we learned?

Understanding the nature of our raw data was crucial to making **informed data cleaning decisions** and identifying **key correlations** that could reveal hidden insights.

Domain knowledge was key to transforming our statistical results into **useful insights** by **contextualizing** and **confirming** correlations present in the data.

Project Update

MVP 1

Model Building:

Identified true null values and developed a preliminary procedure for data cleaning.

Implemented sklearn SVC, KNN, and random forest classifiers to establish a baseline for model performance.

Created a plan for improvements to our data selection and model selection procedures.

Dashboard:

Initial classifications and Key Insights on PowerBI

MVP 2

Model Building:

Improved data cleaning procedure: added imputation and class balancing functions

Further developed KNN and random forest classifier models using hyperparameter grid search and cross validation

Implemented the SHAP explainer to understand the importance of individual variables

Dashboard:

Revamped initial classifications and added data visualization for timeline of failure. Added SHAP visuals and explainer plots

MVP 3

Model Building:

Introduced noise variable and removed collinear variables to perform better feature reduction

Created a function to score the performance of KNN Imputer

Improved SHAP explainer interpretations using affirmations from PCA analysis

Dashboard:

Identifying all variable to validate key influencers of shap models

Brokedown data representing factors infleuncing rod pump failure

Our Procedure and Challenges

Missing Data

Categorical Variables

Non-standardized Null Values

70% Tubing Failures

Data Cleaning,

Feature Selection,

and

Model Training

Hyperparameter Optimization

Collinear Features

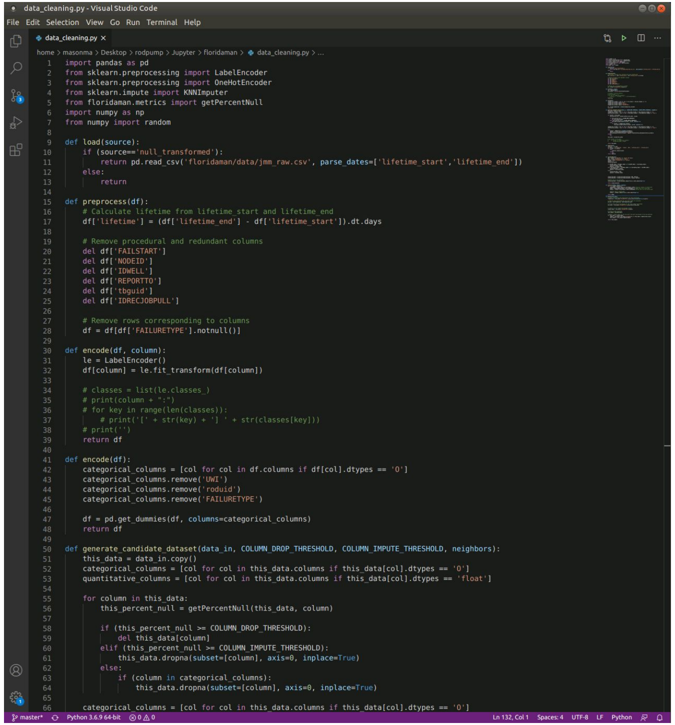
Runtime and Efficiency

Irrelevant Features

Our Procedure

Grid Search:
Data Cleaning Parameters

Our functions:



Standardize Null Values

Address Null Columns:
Drop column or impute

Categorical Variables:
One-Hot Encoder

Uneven Categories:
Balance Dataset

Address Null Columns:
Run KNN imputer

Grid Search:
Hyperparameters

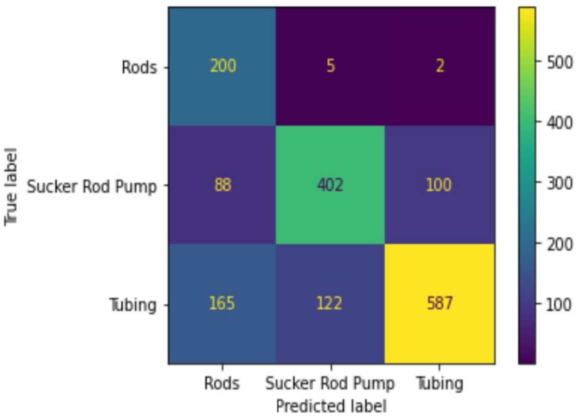
Feature Reduction

SHAP Explainer

Random Forest Model Results:

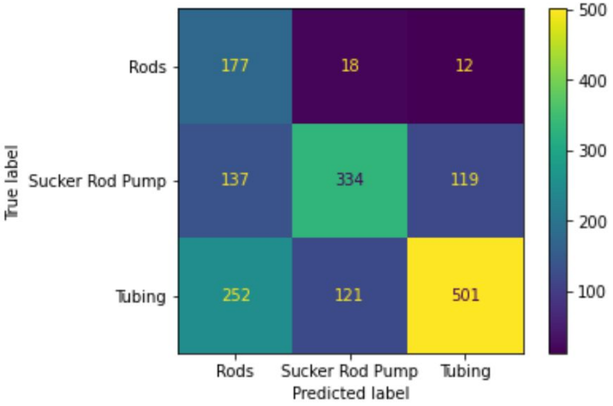
Balancing After Imputation

Accuracy: .71



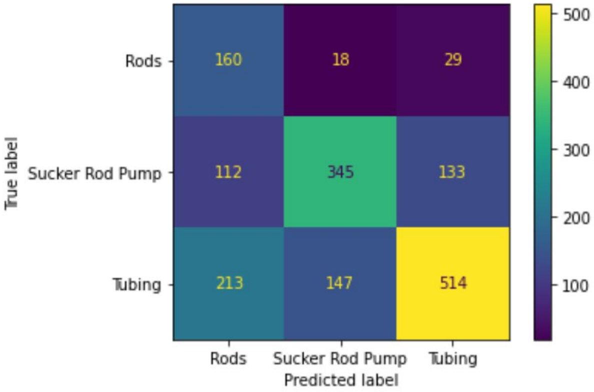
Balancing Before Imputation

Accuracy: .61



Reduced Features

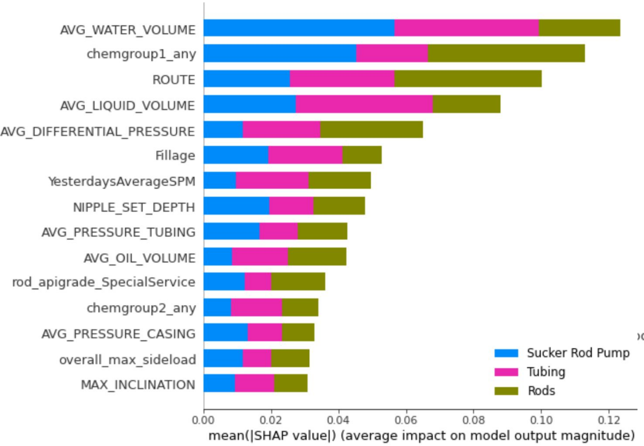
Accuracy: .62



Feature Importance:

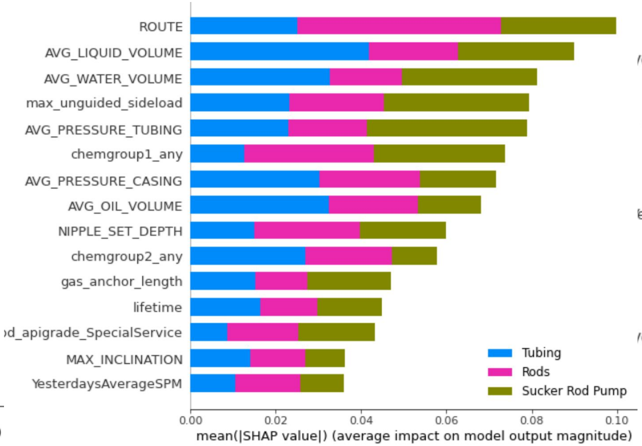
Balancing After Imputation

Accuracy: .71



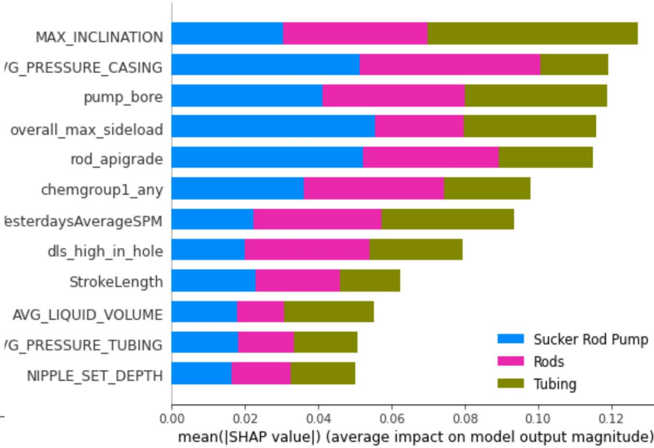
Balancing Before Imputation

Accuracy: .61



Reduced Features

Accuracy: .63



Dashboard and SHAP Update

Live Demo

Business Impact and Recommendations:

Recommendations:

1. When **max inclination** is ...
   a. **> ~2.15,** rods were **8.37%** <u>more</u> likely to classify as a **Tubing** failure.
   b. **> ~2.678,** rods were **13.75%** <u>less</u> likely to classify as a **SRP** failure.
   c. **> ~3.22,** rods were **8.06%** <u>more</u> likely to classify as a **Rod** failure.
2. When **avg_pressure_casing** is ...
   a. **> ~ 112.5** rods were **23.18%** <u>more</u> likely to classify as a **Tubing** failure.
   b. **< ~78.95,** rods were **13.13%** <u>more</u> likely to classify as a **SRP** failure.
   c. **> ~123.8,** rods were **15.01%** <u>less</u> likely to classify as a **Rod** failure.

Business Impact:

1. In light of the above insights regarding failure type classification likelihood, ConocoPhillips can work towards preemptively addressing a rod's failure
2. Help capture majority of lost profit of rod pump failure ($5700/day)
3. Reduce time spent by ConocoPhillips on understanding leading factors contributing to rod pump failure (10 hrs/week)
4. A prepared dashboard and series of shap plots providing data breakdown and likeliness of each failure

Conclusion & Next Steps:

- Improve balancing function and decrease bias introduced by random sampling by implementing a method which selectively chooses samples which well represents the distribution as a whole
- Explore feature decomposition, splitting features into subcategories based on thresholds observed in SHAP values
- Test a one-versus-all methodology which will allow us to train a unique model for each failure type
- New Model Exploration: Gradient Boosted Trees
- Implement Survival Regression Analysis to determine timeline of failures and predict when they will occur