Predicting The Functionality Of Water Wells In Tanzania

Soohyun Kim
Mitch McElderry
Eric Denbin

Presentation Outline

- Purpose of Analysis

- Data & Methods

- Results

- Conclusions

- Future Steps

Purpose Of Analysis

Predict if water pumps are functional or non-functional.

Stakeholder: Tanzanian Ministry of Water

Key Performance Indicator: Functionality

Key Metric: Accuracy

Data & Methods

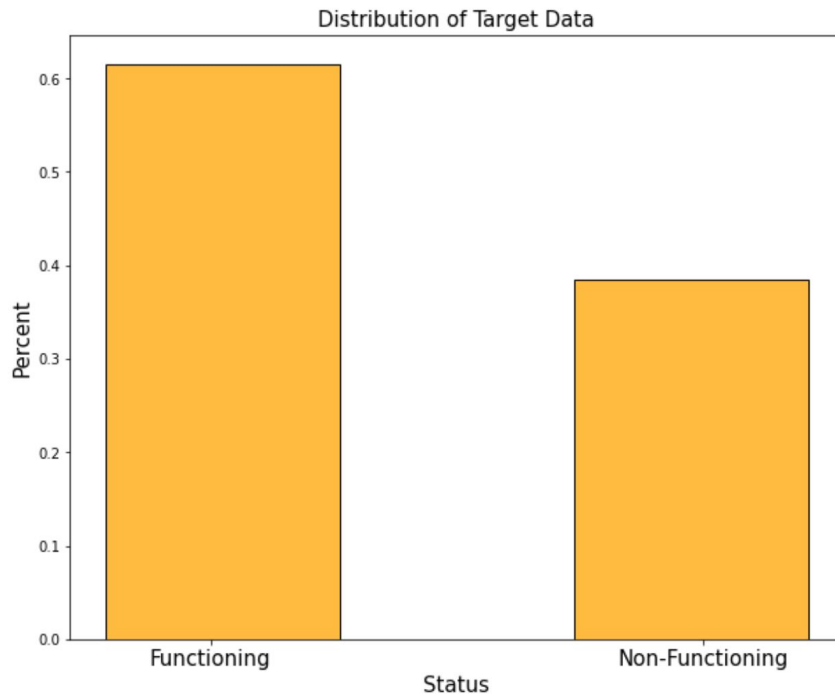**Data provided by DataDriven**: Tanzanian Ministry of Water

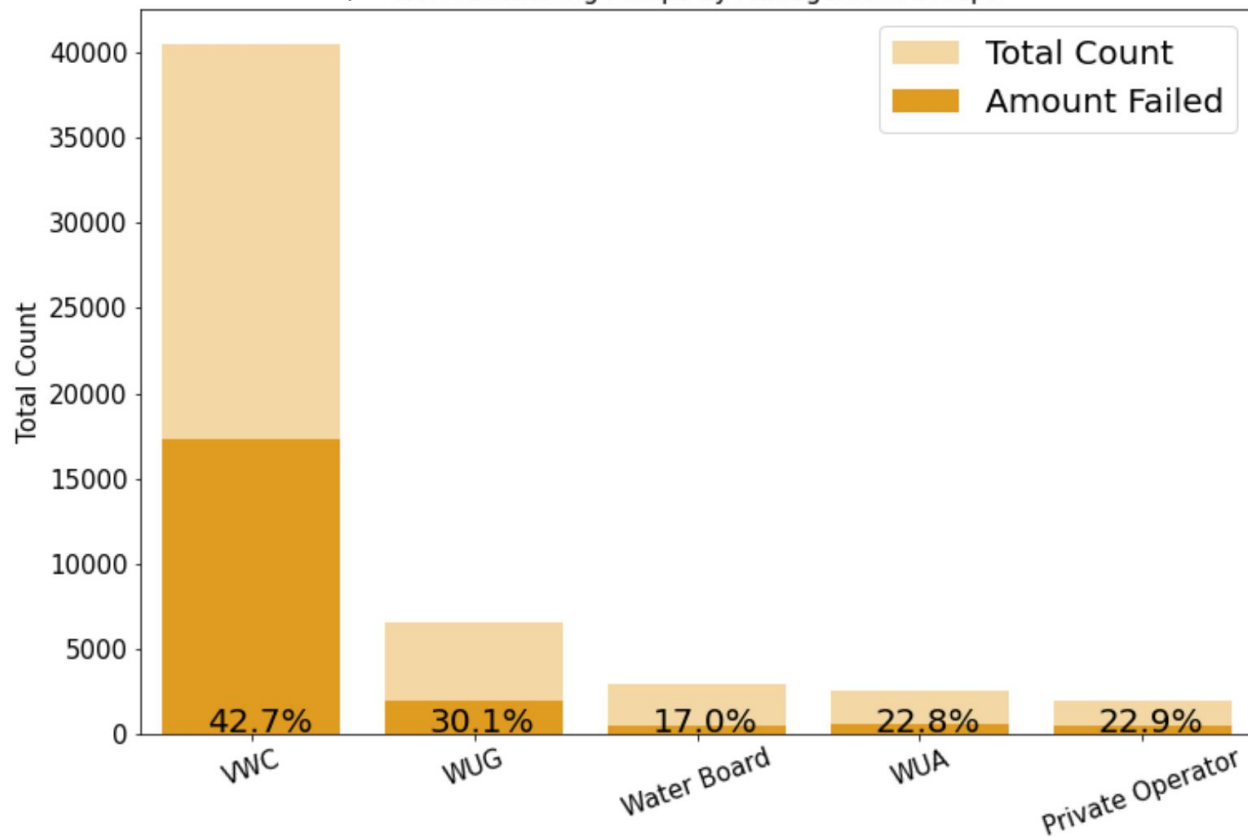**Number of Records:** ~59,400

Important features:

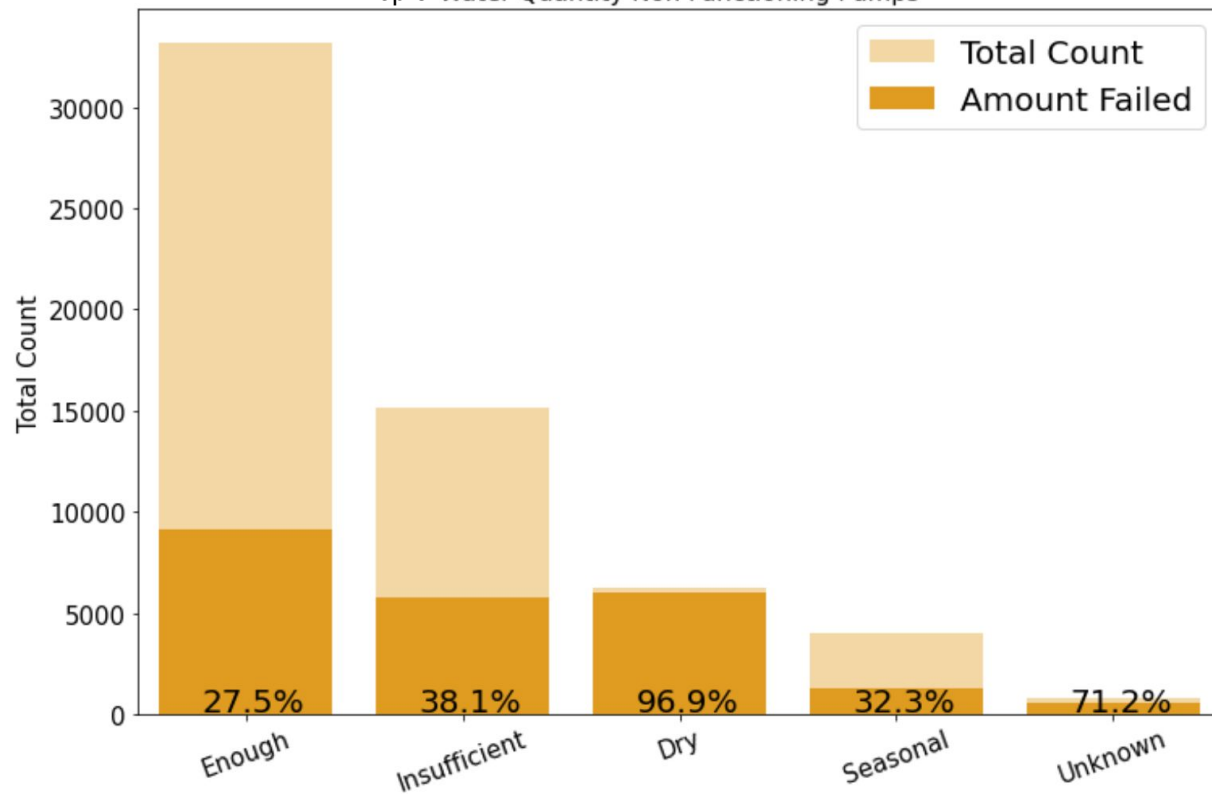- Quantity of water
- Waterpoint Type
- Management Group

Class split: ~60% functional, ~40% non-functional
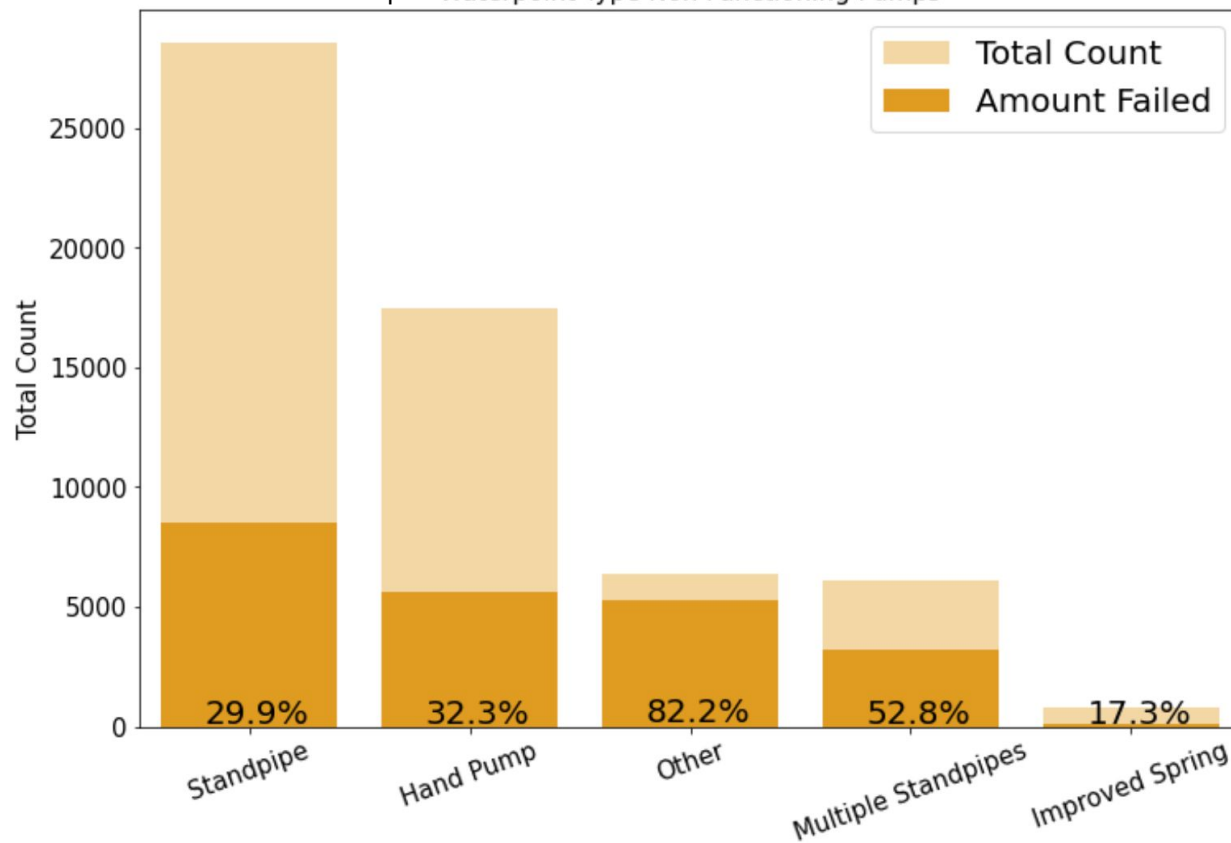

Distribution of Target Data

Top 5 Non Functioning Pumps by Management Groups

Top 5 Water Quantity Non Functioning Pumps
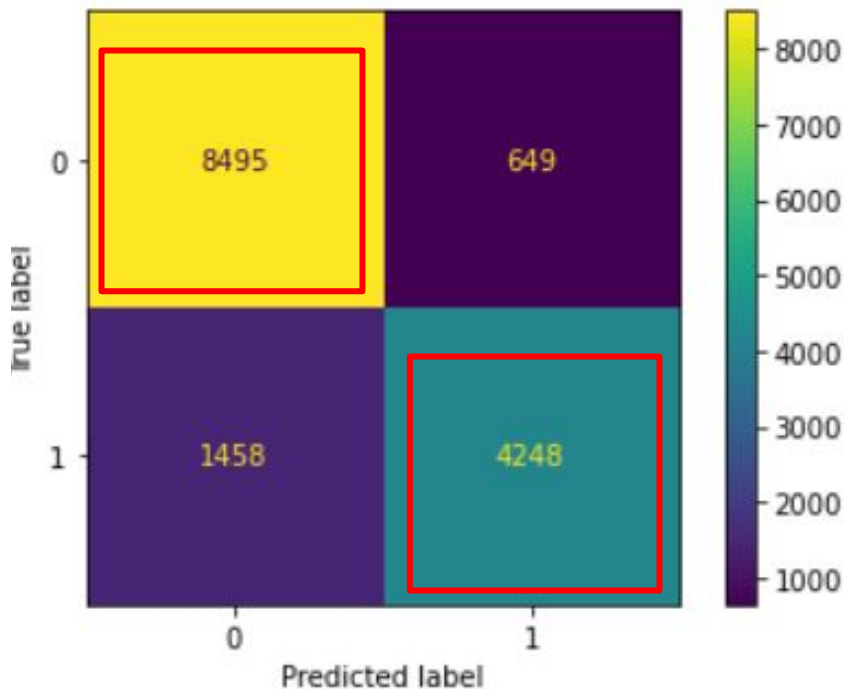
Top 5 Waterpoint Type Non Functioning Pumps
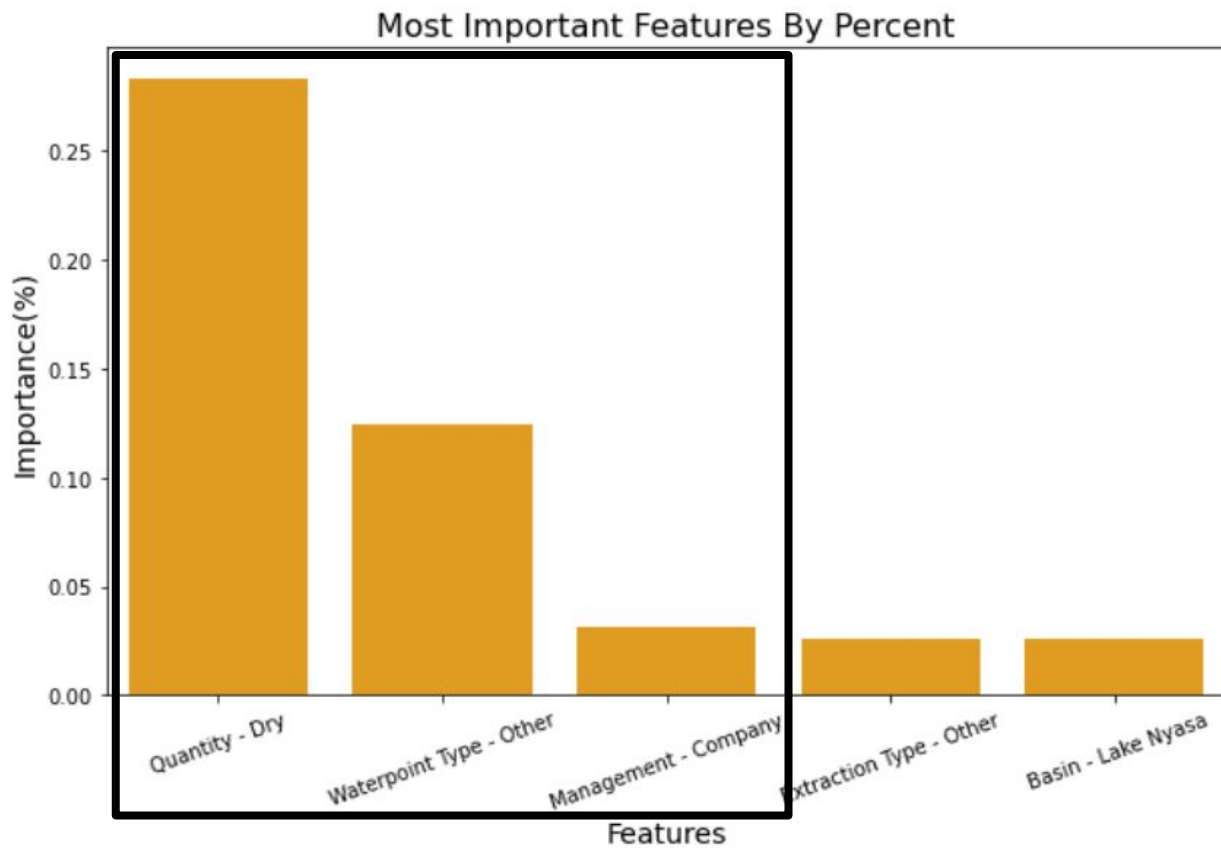
Results - XGBoost Model

Final Model Accuracy:

85.8%

Results

Most Important Features By Percent

Recommendations

- Perform regular checks of water quantity for all pumps

- Replace 'other' waterpoint types with gravity waterpoint types

- Put policies in place that ensure the management company of a well are performing regular checks and maintenance

Future Steps

- Identify the categories of certain features that are labeled as 'other' and 'unknown'.

- Use machine learning to create a model that predicts the quantity of water in a well

- Determine if the funders of a well impact its functionality

Thank you

Eric Denbin - Github @ericdnbn

Soohyun Kim - Github @soohyunnie

Mitch McElderry - Github @mmceld2