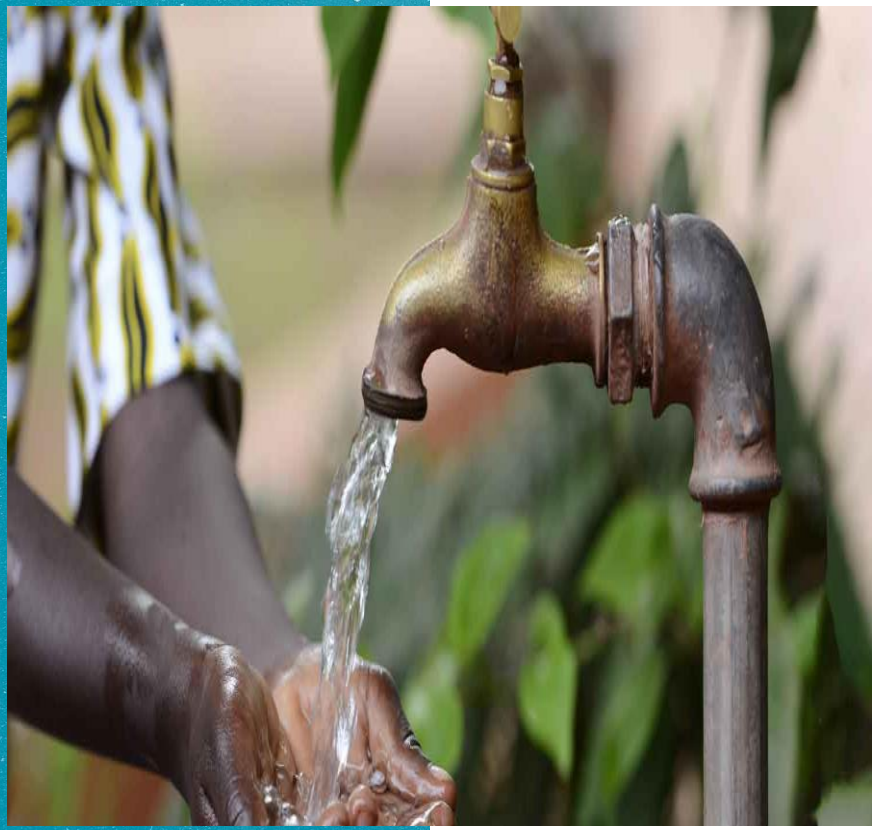
Tanzani Water Wells Classification

Introduction

STAKEHolders:
Taarifa and **The Tanzania Ministry of Water.**

Goal:
To aid in Predictive Maintenance of water wells
Using Machine Learning

• Problem Overview



Access

- While much of Tanzanians population has access to basic water services, a large 39% of households still lack this basic need

Maintenance

- Predictive maintenance of the water wells that provide water to much of the rural population is needed.

Objectives

Feature Identification

To identify the best features to use in building the predictive model

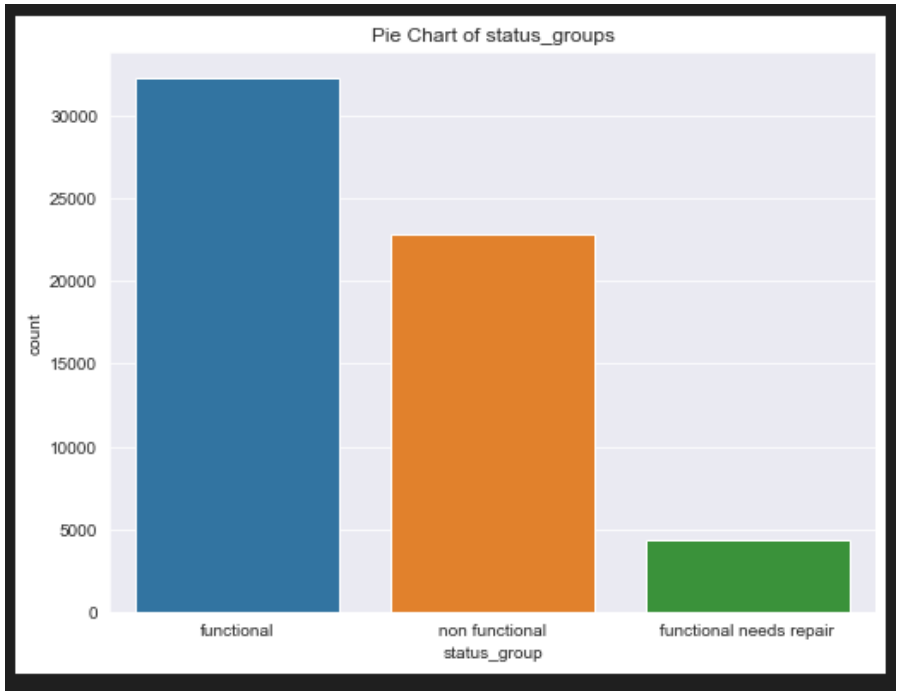
Modeling

To build and train a classification model that can predict which wells are likely to need repairs.

Evaluation

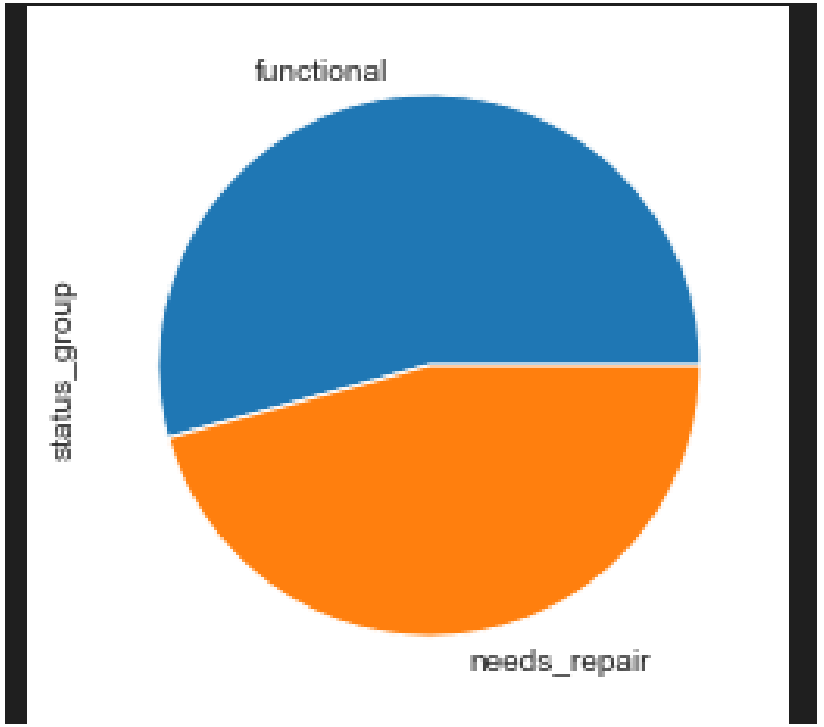
To evaluate the effectiveness and improve performance of the model built.

Feature Identification



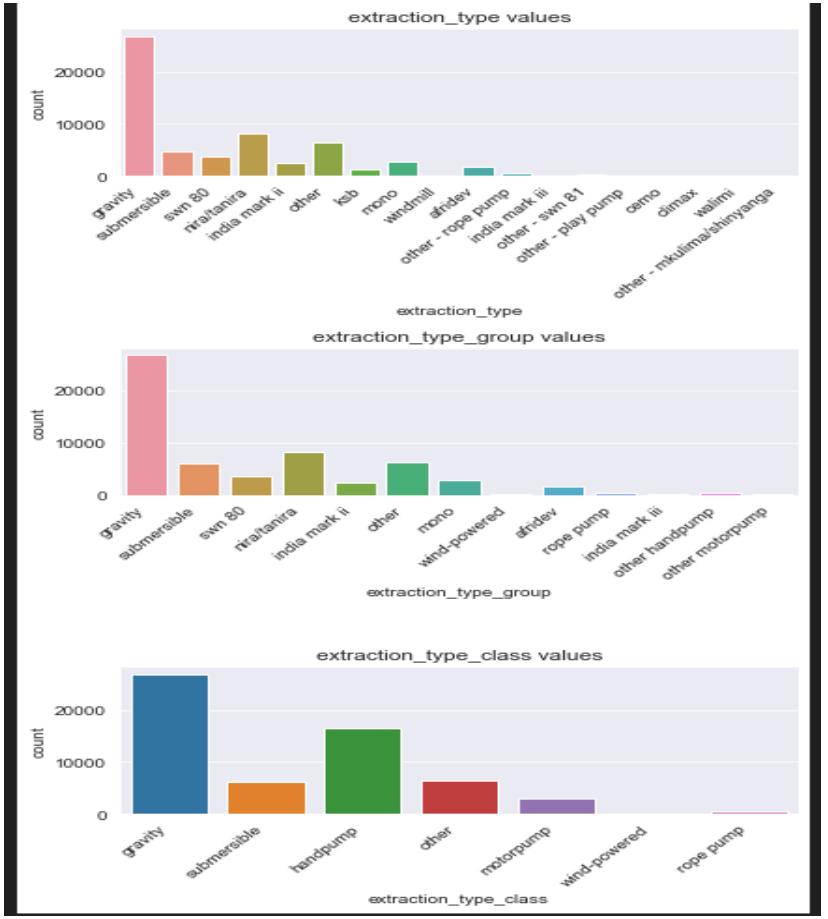
The distribution of wells by the functioning status.

Feature Identification



To reduce class imbalance we combined the two columns into one called ‘needs repair’

Feature Identification



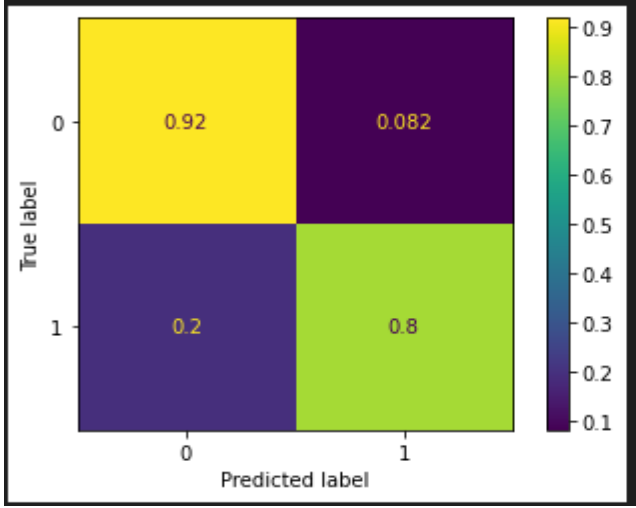
To better generalize some, some columns were dropped in the model.

Modelling

```
[0.87017957 0.87073562 0.86835219 0.87018382 0.8701184 ]
Train Accuracy 0.8699139197696514


[0.79173208 0.79029304 0.80321863 0.7960225  0.78437786]
Cross-Validation Accuracy 0.7931288231030292


Training Recall: 0.808793825525506
Test Recall: 0.7180544375746509
```



The best model was a Logistic Regression model that was able to predict with 79% accuracy.

Evaluation

Best Model:
Logistic Regression
Accuracy:
79%

Recommendations

- More data can be sourced to improve the model
- The model can be deployed in an environment of choice to help ensure water security.

Thank you