
Tanzanian Water Wells project: Status Report

David Wabuko. [Data Scientist]

Overview

Expected delivery

October 2025

Recent progress

- The Project is meant to carry out analysis on Water wells
- The Analysis is to know the water wells that require Repair in Tanzania

Biggest risk

The Biggest risk is incorrectly predicting water wells that require repair

Data Analysis

Analysing Data

After thorough analysis the Data is modelled and Data science methodologies applied.

Data Modelling

- This will entail training data and coming up with models that will ensure we get relevant insights.
-

Models and Methods

Decision Tree Model

- I came up with a decision tree model for our data training it and testing it
- This Model came up with the best results after comparing Metrics and Tuning

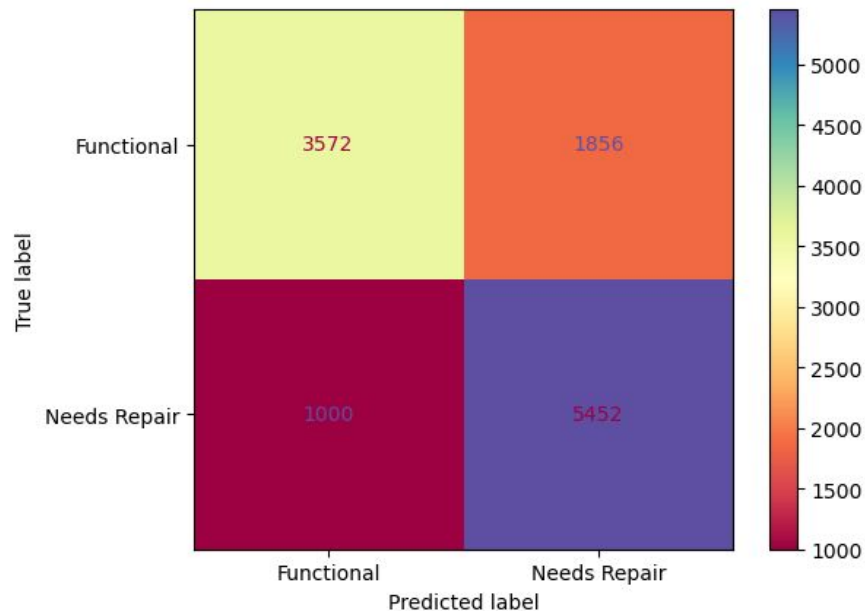
Logistic Regression Model

- This model gave results but not results we can rely on to Predict our waterwells, This was a struggling model for the nature of this data.
-

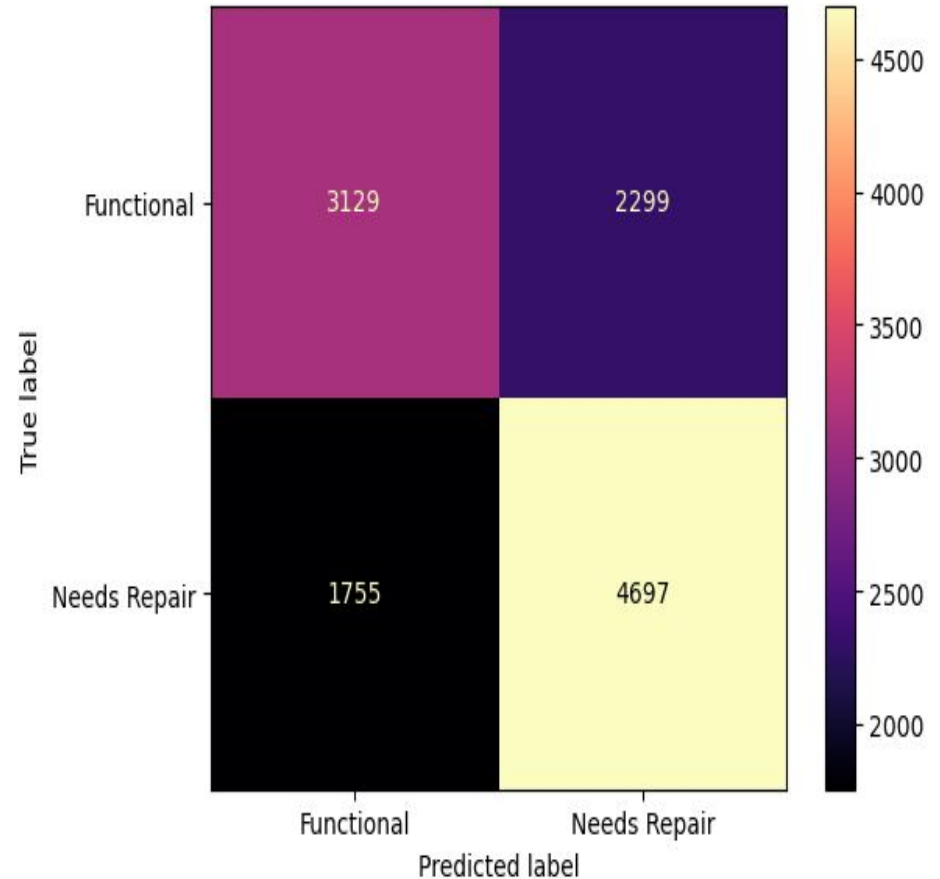
Model Performance

Our Decision Tree model performed well on both our Training and Test sets giving exemplary results.

Results are best explained in the technical Notebook.



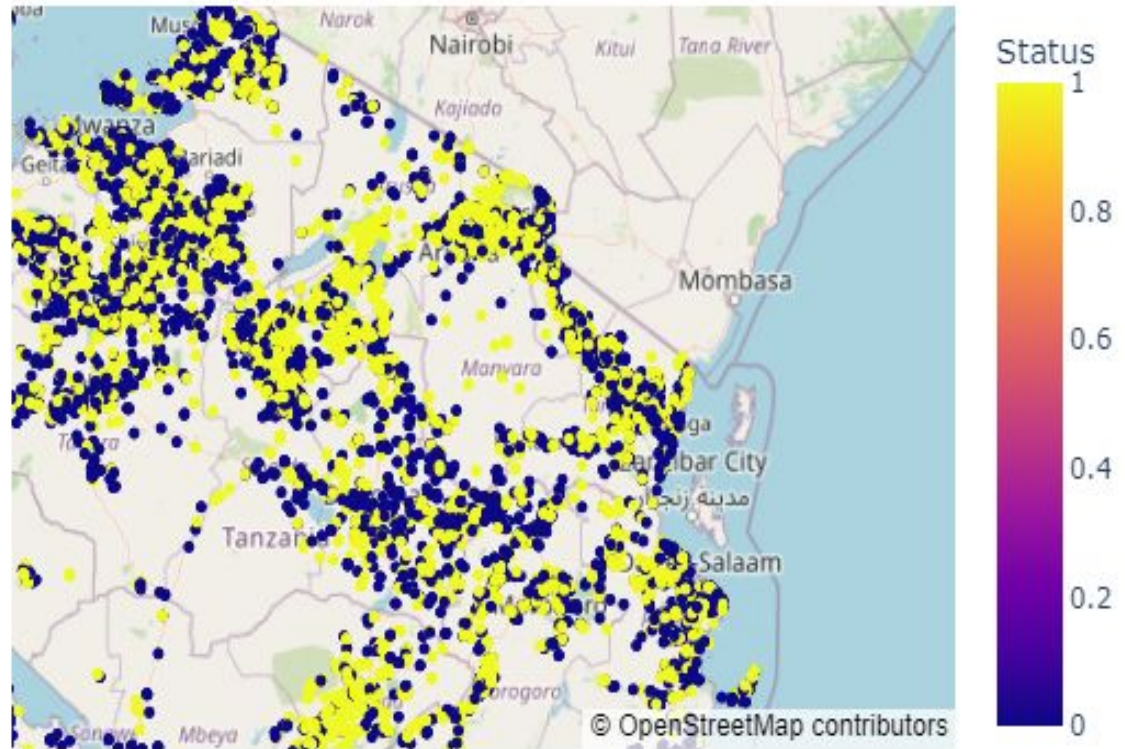
The Regression model which showed signs of Overfitting and Underfitting had results but not as good as the decision Tree. even after refining.



Model Deployment

I Deployed the model I chose and it is working perfectly
The results of Non- Functional and Functional waterwells is
precisely classified by longitude and latitude hence giving access to
correct coordinates and exact location of the Waterwells. Results
were displayed in a Map.

Status of Waterwells



1. I recommend my Decision Tree Model For its preciseness in locating Waterwells Location and Functionality Status of the Pumps within the wells
2. Through knowledge in this area I recommend the use of Non-submitted Pumps because they break down faster, a factor caused by a drop in level of Water in the well hence they tend to suck up and pump mud causing clogging issues.
3. The most Probable cause of failing Pumps is as a result of Poor maintainance, I would recommend reshuffling the maintainance groups or deploying new maintainance groups for the Pu

