Tanzania Water Pumps Status

By:

EDWIN MAINA

- Over 24 million people are impacted by the The United Republic of Tanzania's water crisis; that's almost half of the population of Tanzania.

Business Understanding

- Access to clean water is a fundamental human need and a key for public health, economic development, and social well-being.

- In Tanzania, communities heavily rely on water wells for their daily water supply. However, many of these water points become non-functional over time due to poor maintenance, environmental conditions, or inadequate infrastructure.

- This project aims to build a predictive models to assess the functionality status of water wells in Tanzania. The models will help stakeholders prioritize maintenance efforts, allocate resources more effectively, and ensure that water wells remain operational for the communities that depend on them.

Business Problem

Tanzania's water supply system is characterized by frequent water pump breakdowns resulting from lack of proper maintenance and inefficient management. This leads to disruptions in water supply, exacerbating the acute shortage of clean water and socio-economic losses.

Stakeholders

- Government of Tanzania: Interested in optimizing resources for maintaining and repairing water wells to ensure clean water access for citizens.
- NGOs focusing on clean water access: Seeking to identify and prioritize wells in need of repair to efficiently allocate resources and interventions.
- Private sector partners: Potentially interested in investing in water infrastructure projects a

Objective

The objective of this project is to predict the functionality status of water pumps based on various features related to the pumps' installation.

- Evaluate factors that affect the functionality of a pump.
- Identify and model combinations of features that best predict the functionality of a water pump.
- Test and validate the accuracy of the model.
- Draw conclusions and recommendations.

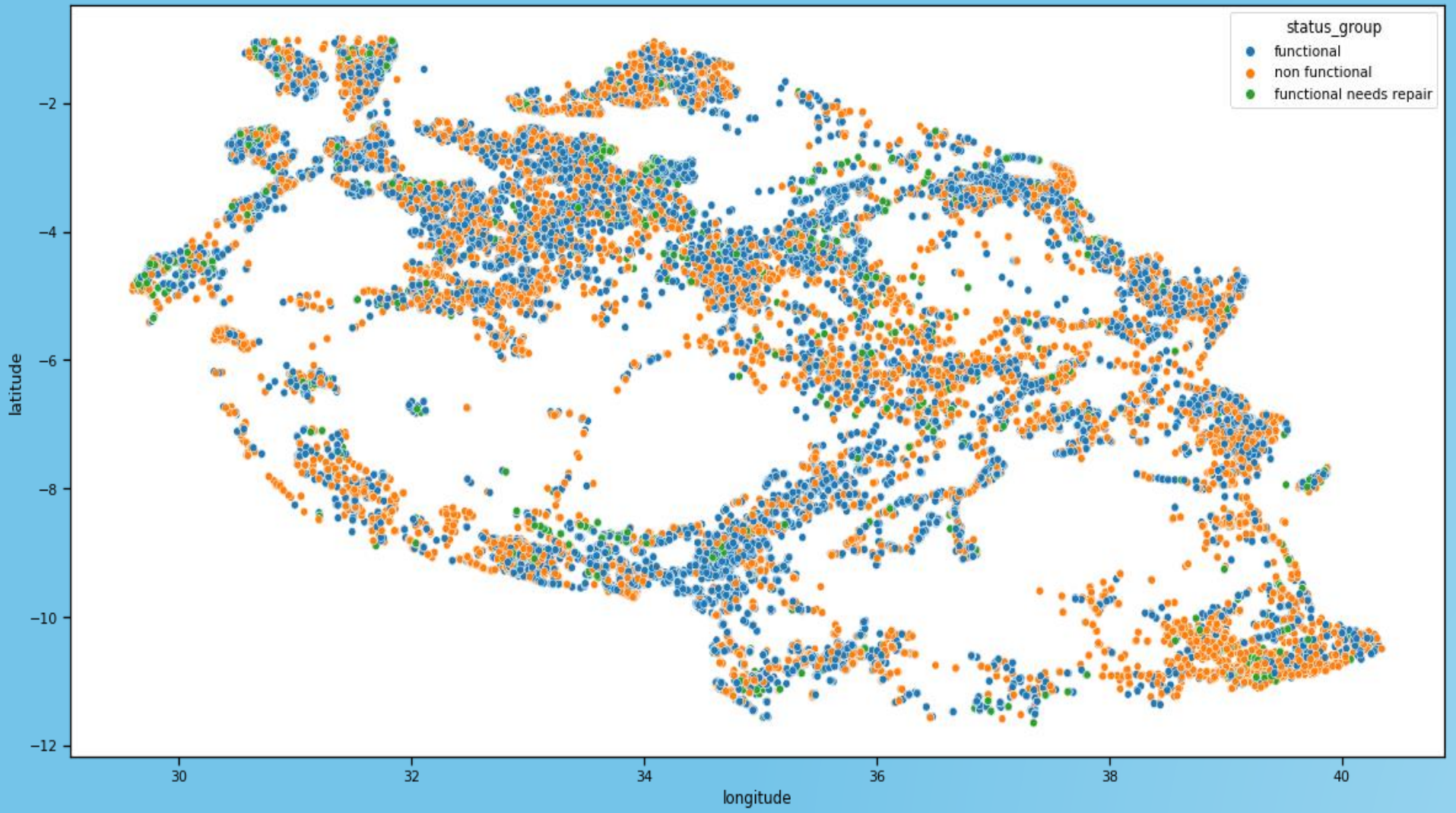
DATA

The data used in this project is from the Pump it Up: Data Mining the Water Table competition hosted by DrivenData, originally sourced by Taarifa and the Tanzanian Ministry of Water

Data Preprocessing

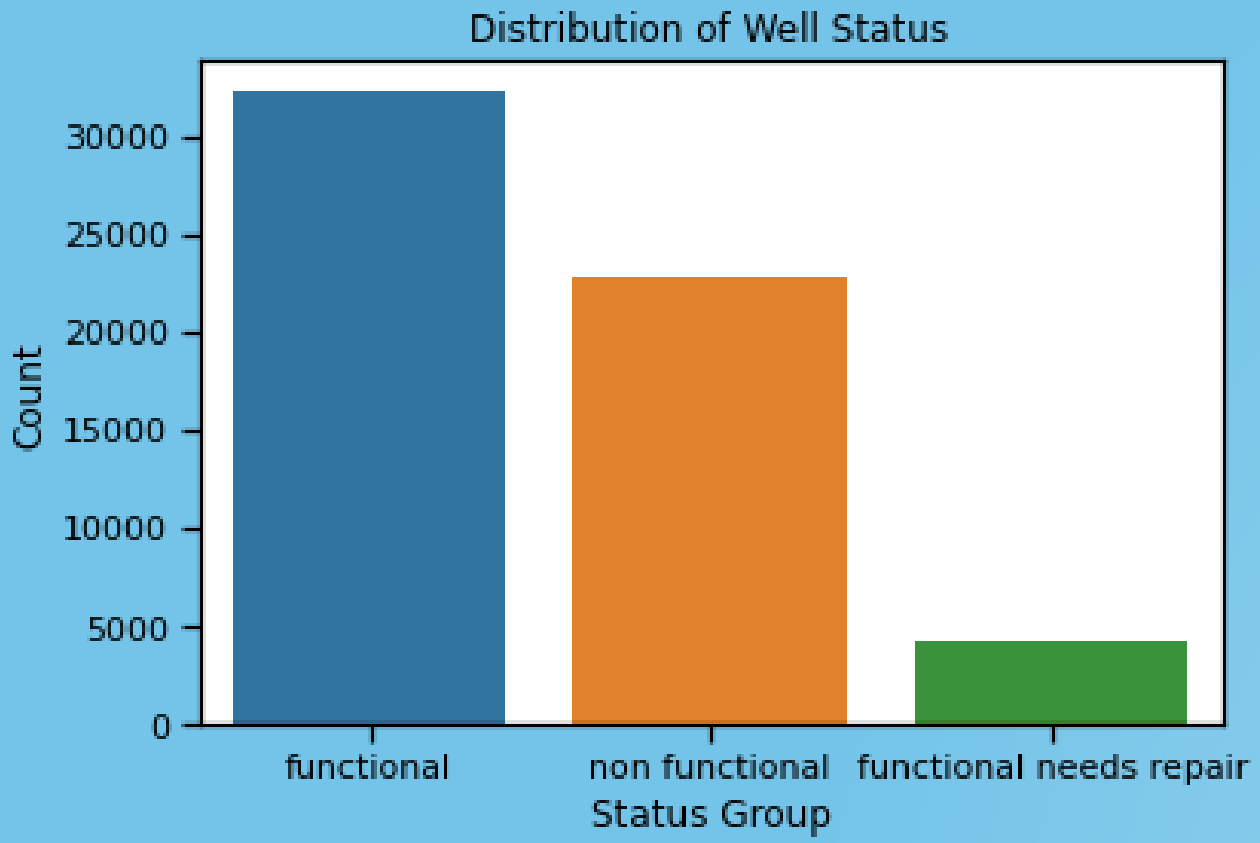
- Data Cleaning: I made sure our water pump data was complete by filling in any missing information, especially for who funded the pump and who installed it.
- Feature Engineering: figured out each pump's age from when it was built to help us understand which pumps might need more attention.
- Encoding: organized the pumps' details, like water quality and location, into clear categories so our analysis can be more accurate.
- Class balancing: adjusted the numbers to make sure no single detail about the pumps would be more important than another in our study

Exploratory Data Analysis



The functionality status of water pumps appears evenly distributed. However, the Southeastern and Western regions demonstrate a significant concentration of non-functional water pumps.

Status Group



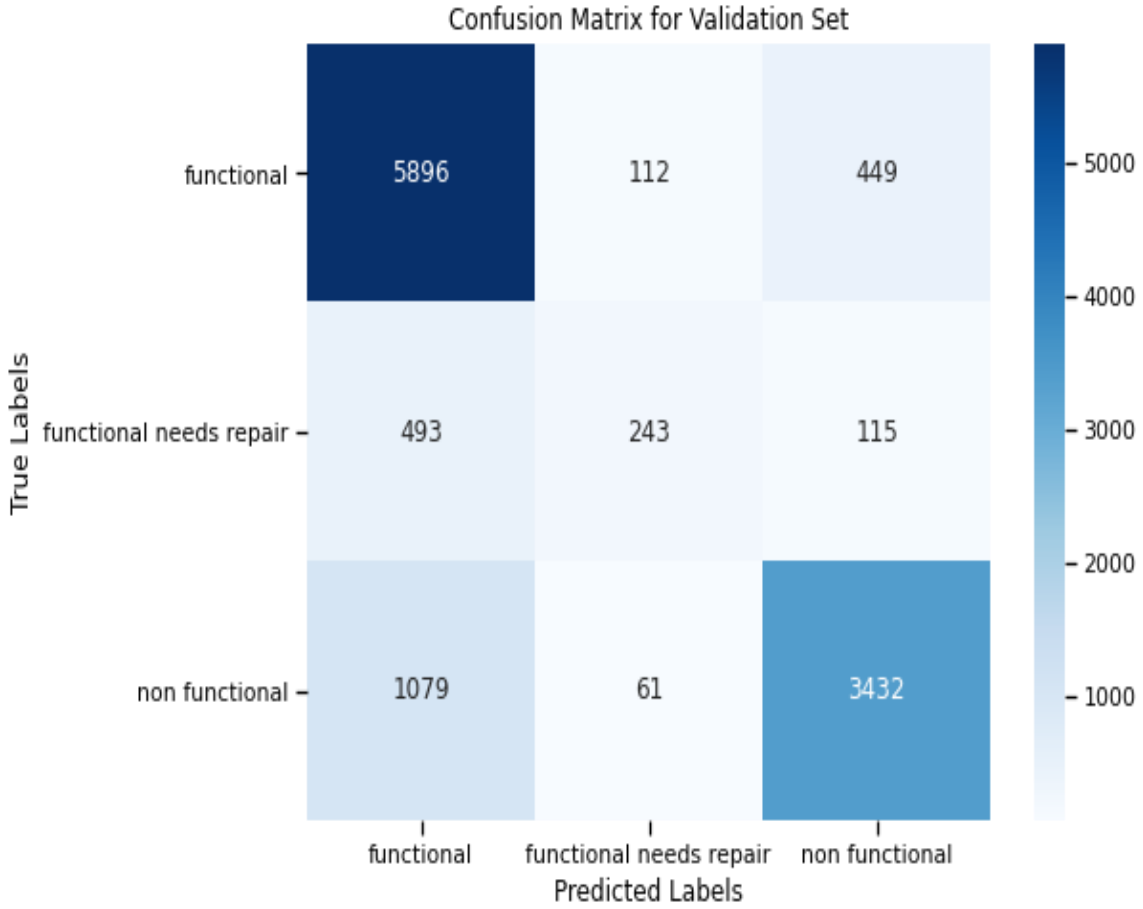
Most of the wells are in a functional status but there is a significant number of water pumps that are non-functional

```
Tuned Random Forest Best Hyperparameters:
max_depth: 20
min_samples_leaf: 1
min_samples_split: 2
n_estimators: 500
```

	Model	Train Accuracy (%)	Validation Accuracy (%)
0	Logistic Regression	72.92	72.94
1	Decision Tree	99.54	75.87
2	Random Forest	99.53	80.08

- **Modelling**

- Logistic regression, Decision trees and random forest models were evaluated and the best performing model Random forest had highest accuracy of 80.08 and was picked for tuning with Gridsearch was to be the final model



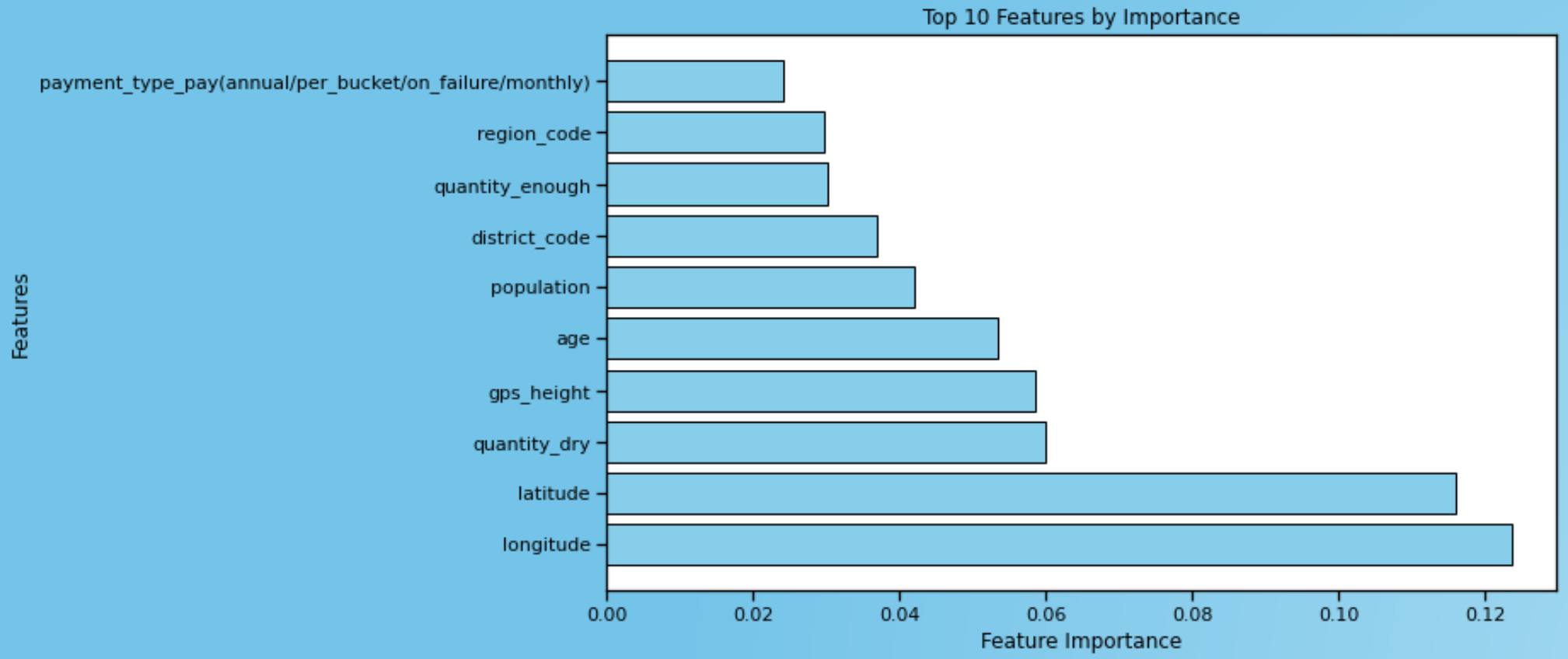
- Analysis
- Accuracy: 80.56%

- The model achieved an accuracy of about 80.5%, indicating a fairly good performance in classifying the instances correctly
- F1-score across classes is moderate, as it doesn't consider class imbalance.
- Weighted Avg F1-Score: 0.80 The weighted F1-score reflects the model's better performance in more frequently predicted classes (Functional and Non-functional).

- The most important features according to the final model are longitude, latitude, quantity_dry , age, region code and population.

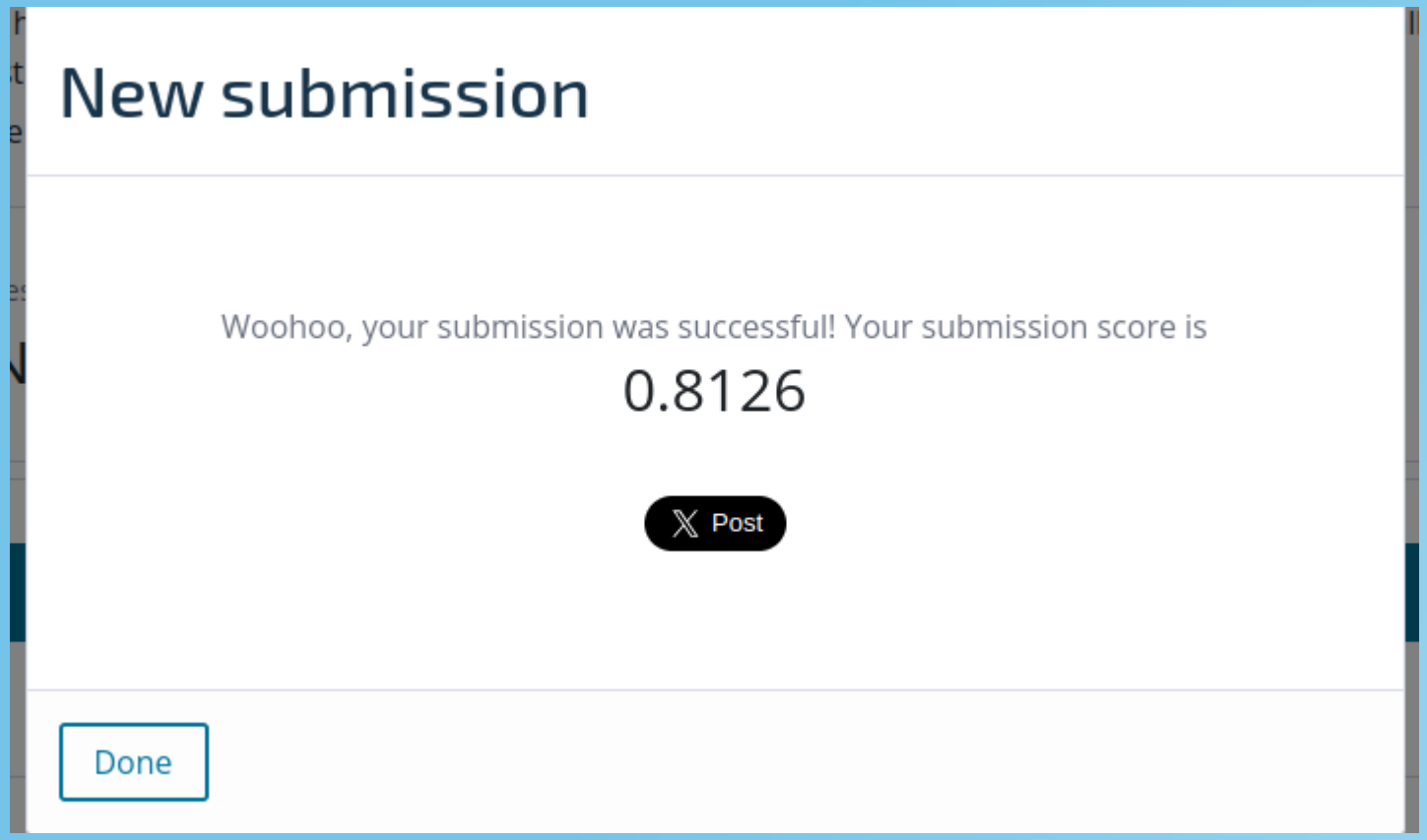
Limitations

The overall data used was not up to date
The data was not reliably gathered, it contained a lot of placeholders in important features such as population and amount_tsh, hence the results obtained are not particularly accurate.



Conclusions

• Machine learning algorithms prove to be effective in predicting the status of waterpoints in Tanzania with notable accuracy.
• Regression models show significant promise, highlighting the predictive potential for waterpoint conditions with accuracies reaching over 79%.



RECOMMENDATIONS

- Work with the local government to ensure more accurate gathering of data

- Data collected should highlight more on non-functional pumps or those in need of repairs

- Financial Barriers:
- Recognize that paying for maintenance services poses a significant challenge, given the average income levels in Tanzania. Introduce financial support programs at the local government level and explore sustainable payment models to ensure continued waterpoint maintenance.

- Continuous Data Collection
- efforts to gather consistent, high-quality data to improve predictive models and enable preemptive waterpoint maintenance.

- . Pump Improvement - Research and deploy efficient pumps.

THANK YOU

QUSETIONS