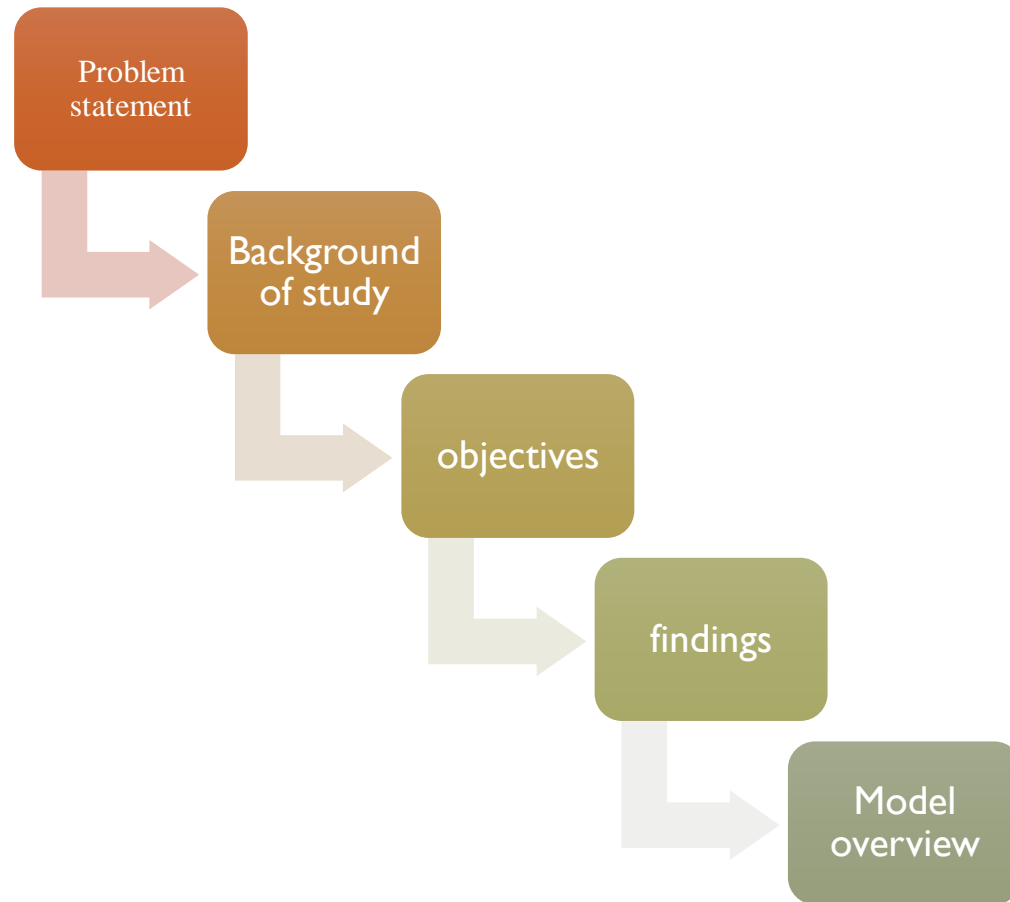




TANZANIA WATERPOINTS ANALYSIS AND SOLUTIONS.

Predicting the functionality of water wells in Tanzania using
machine learning.

OVERVIEW



PROBLEM STATEMENT

The Ministry of Water in Tanzania needs to ensure that clean, potable water is available to Tanzania's inhabitants.

The water problem can be easily solved with knowledge of the functionality of the different waterpoints.

Classifying the functionality of a water point to be either functional, functional but needs repair or non functional, assist the government know the waterpoints that either need improvement, or those that are functional.

If we can accurately classify a waterpoint, the Ministry will have a better understanding of their existing infrastructure, and because of cost savings, will be able to reallocate existing resources to expand the water infrastructure.

BACKGROUND OF STUDY

Poverty Rate in Tanzania

14 billion inhabitants of Tanzania live under extreme poverty conditions with an average of \$ 1.9 dollar living expenditure per day

Clean water requirement and sanitation

4 million people in Tanzania lack access to an improved source of safe water, and 29 million don't have access to improved sanitation.

It is estimated that Tanzania spends 70 per cent of its health budget on preventable WASH-related diseases as the majority of the population does not have access to improved sanitation, and close to half of the population does not have access to clean drinking water



ULTIMATE OBJECTIVE

Create solutions to the water crisis problem by predicting the functionality of the water pumps whether they are functional, needs repair or completely non-functional.



RESEARCH FINDINGS

More than 60000 thousand wells are distributed within Tanzania.



Functional water wells

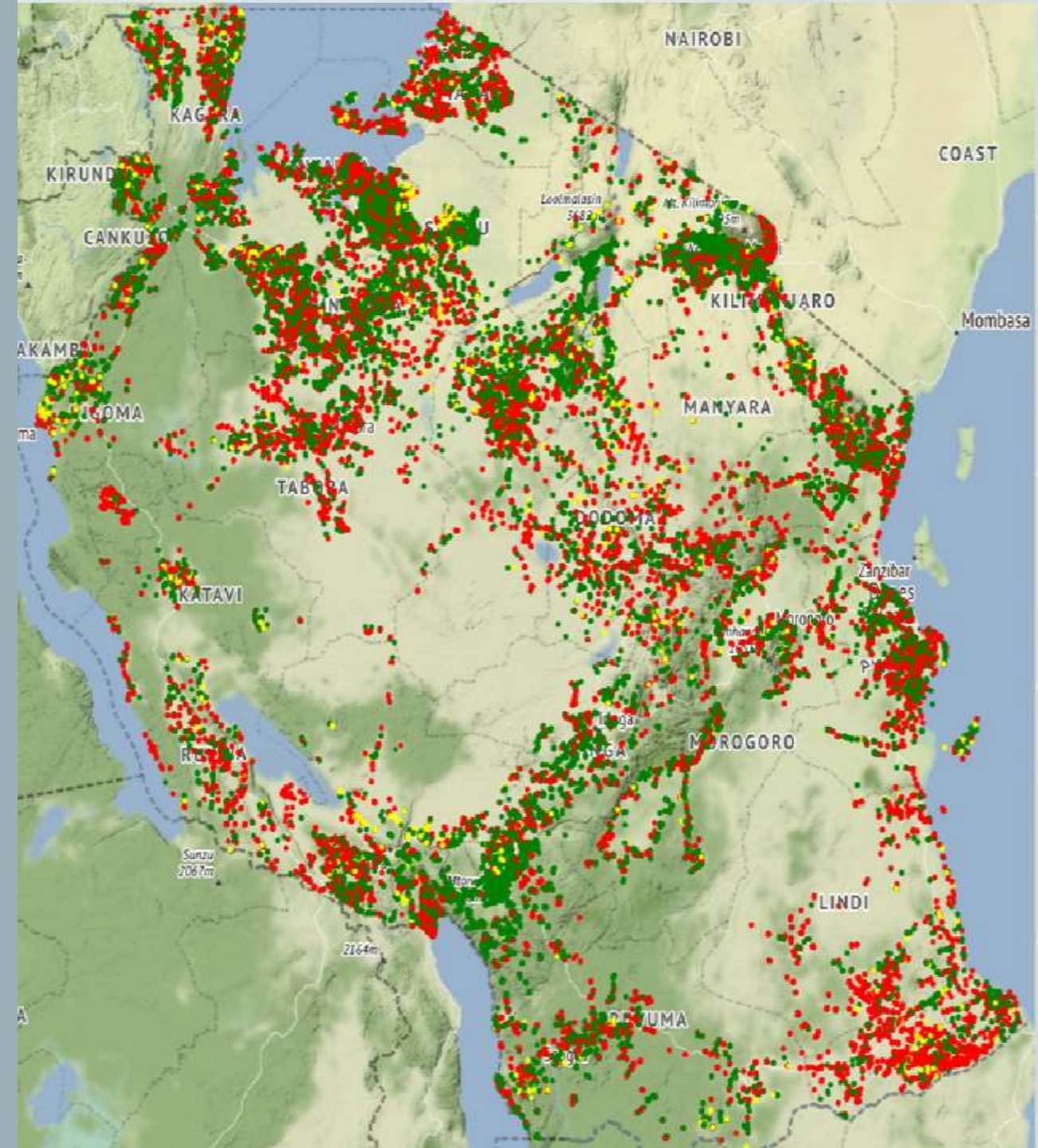


Functional but needs repair



Non functional

Using the Taarifa dataset features, the functionality of the water points can be predicted.



INSIGHTS

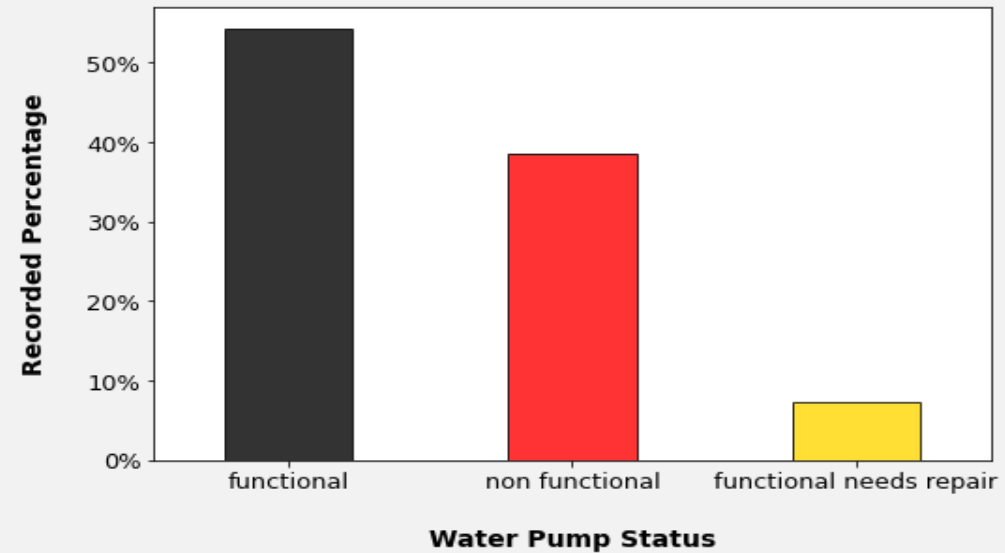
Important prediction features from the Taarifa dataset

FUNCTIONALITY DISTRIBUTION

DISTRIBUTION TARGET VARIABLE



Water Pumps Functionality Status Spread



Functional water points from more than 50 % of the waterpoints.

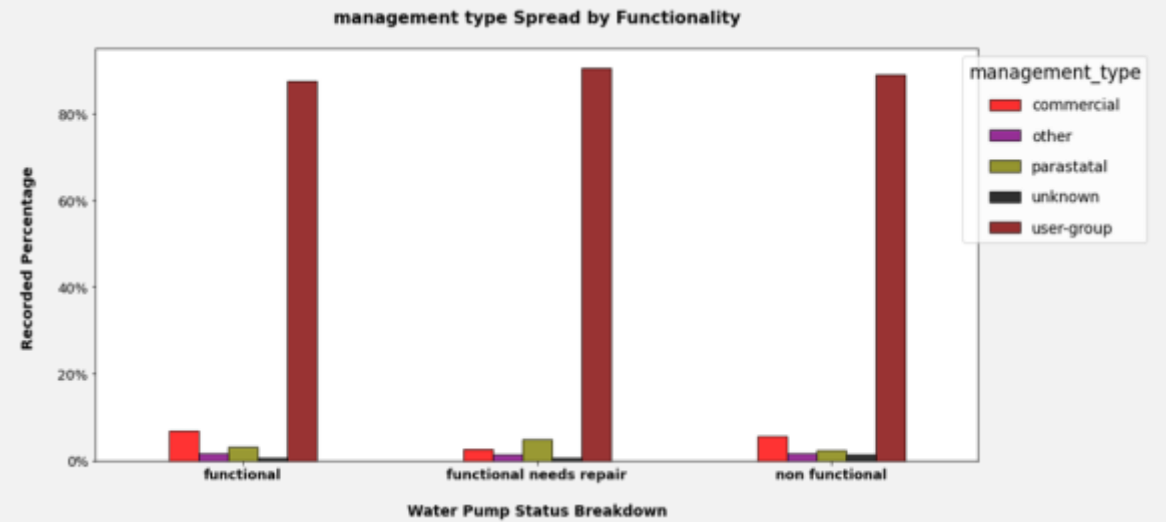
Waterpoints that need repair though are still functional form less than 10%.

Non functional water points are more than 35% of the waterpoints.

WATERPOINT MANAGEMENT



MANAGEMENT DISTRIBUTION BY FUNCTIONALITY

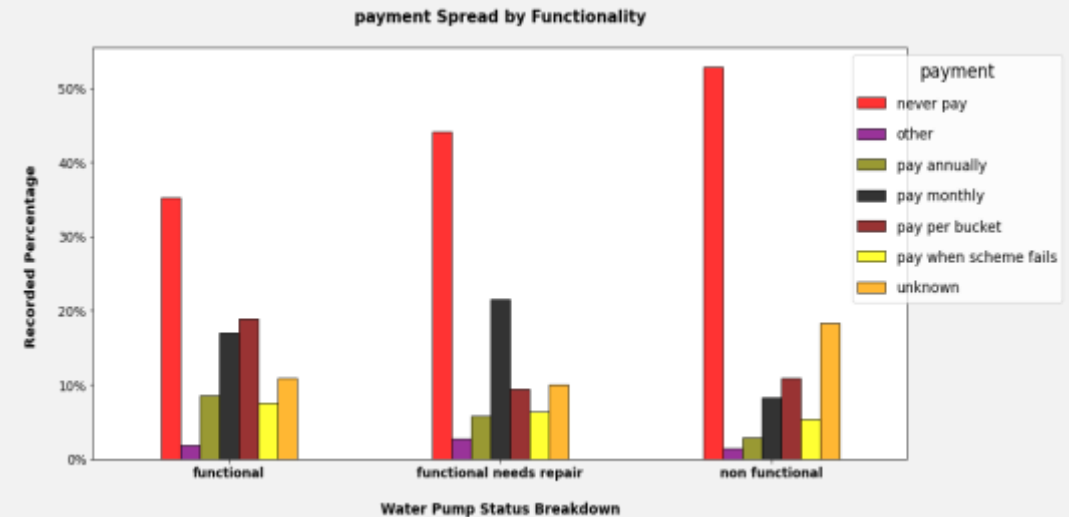


- More than 80 percent of the waterpoints are managed commercially.
- The highest category of the functional, non-functional waterwalls are managed commercially.

PAYMENT TYPE



PAYMENT TYPE AND FUNCTIONALITY



More than 50 percent of the non functional water points are not paid for.

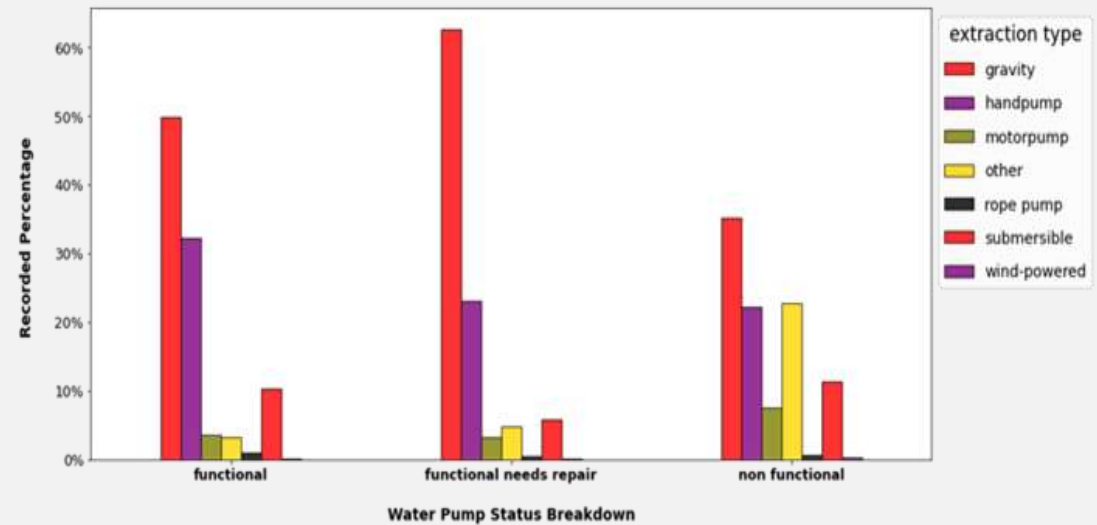
The least percentage of non functional water points are paid for per bucket (clear illustration of the poverty in Tanzania)

EXTRACTION TYPE



DISTRIBUTION OF EXTRACTION TYPE WITH FUNCTIONALITY

extraction type Spread by Functionality



Gravity is the most common type of extraction.
Has the highest percentage of waterpoints that need repair.

RECOMMENDATIONS



- targeting repairs to clusters of wells especially those with high populations
- payments of some kind will provide an incentive to keep wells functional
- allocate funds and resources to effective organizations with a track record
- Use the model to prioritize site visits. Priority should be given to maintenance staff sent to waterpoints that are predicted to be functional but in need of repair or are non functional.
- The ministry can use the analysis as a tool to choose installer with the most long lasting pumps by looking at the age of the well and its functionality.
- By looking at the predicted values, the ministry can figure out the management group with the highest functional pumps and extraction type that when used are the most functional
- The Ministry can use the accuracy of the model and its improved maintenance program as a selling point when soliciting international aid.

MACHINE LEARNING MODEL



The cross validation score for this model is 89%. This score shows that the performance of the model is standard since it can be able to predict 89% of the test data.

```
log loss: 0.4106744518960779
accuracy: 0.8078563411896745
precision: 0.7956785443517816
recall: 0.8687913907284768
```

FUTURE IMPROVEMENTS

Incorporate other factors like the climatic conditions

Update model regularly for maintenance purposes.





THANK YOU

Presenter: Faith Leo



Questions?