Predictive Maintenance of Tanzanian Water Pumps

A proposal to the Ministry of Water to use a Machine Learning Classifier to Improve Resource Management

Outline

1

Business Problem

What is the current situation and how could it be improved?

2

Data

What is available to feed our classifier?

3

Model Considerations

What makes a good model?

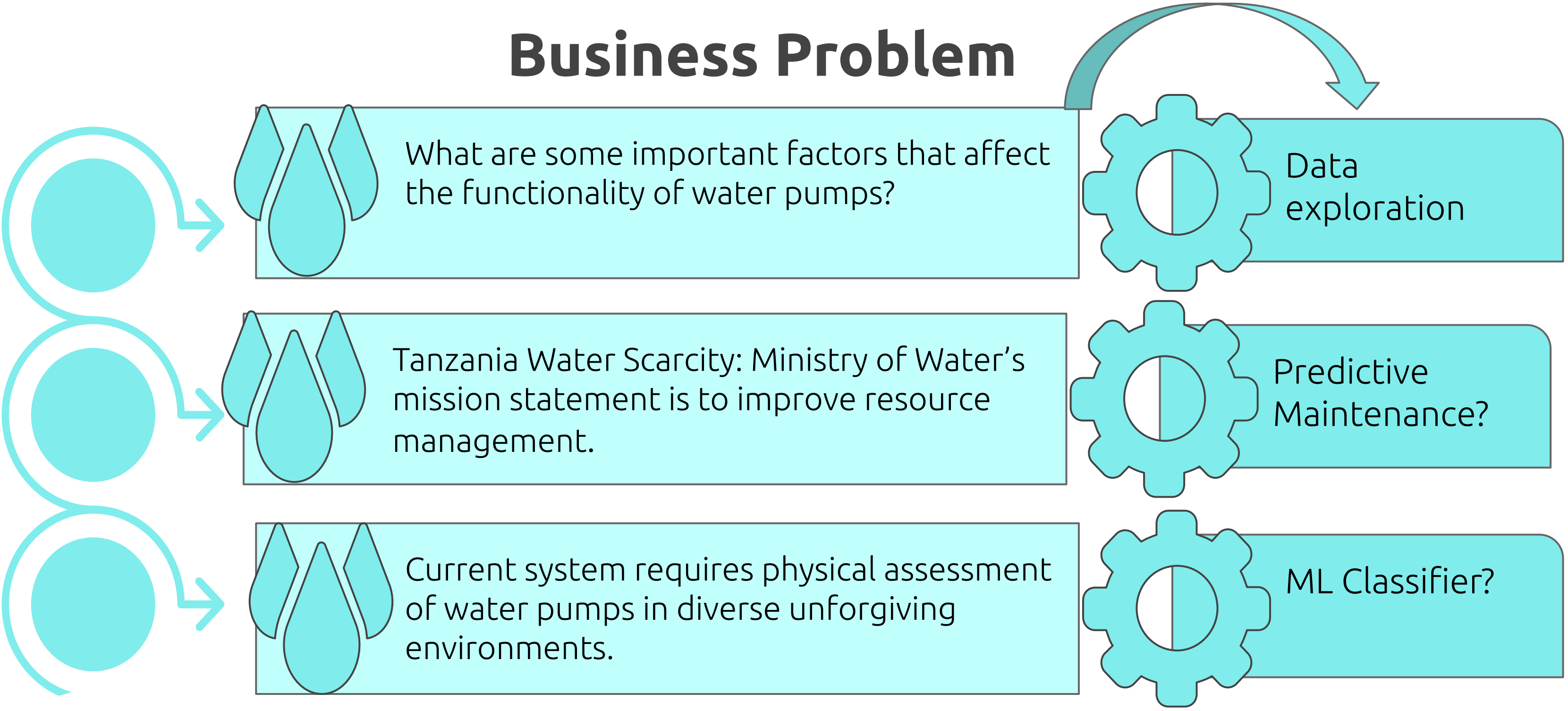
Outline

4 **Model Evaluation**
Which model is the winner?

5 **Conclusion**
What do we know now?

6 **Future Work**
Next steps to make improvements.

Business Problem



What are some important factors that affect the functionality of water pumps?

Tanzania Water Scarcity: Ministry of Water’s mission statement is to improve resource management.

Current system requires physical assessment of water pumps in diverse unforgiving environments.

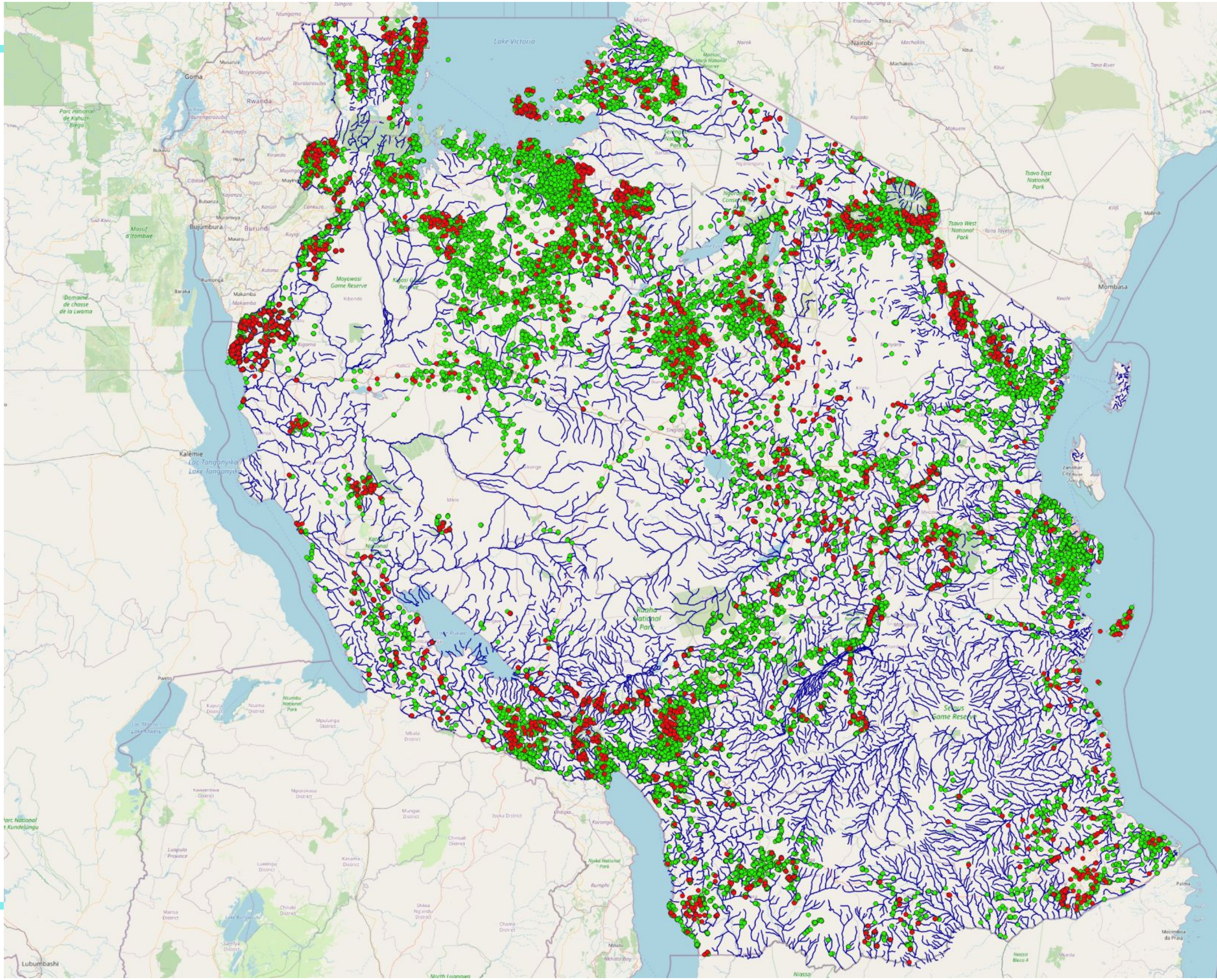
Data exploration

Predictive Maintenance?

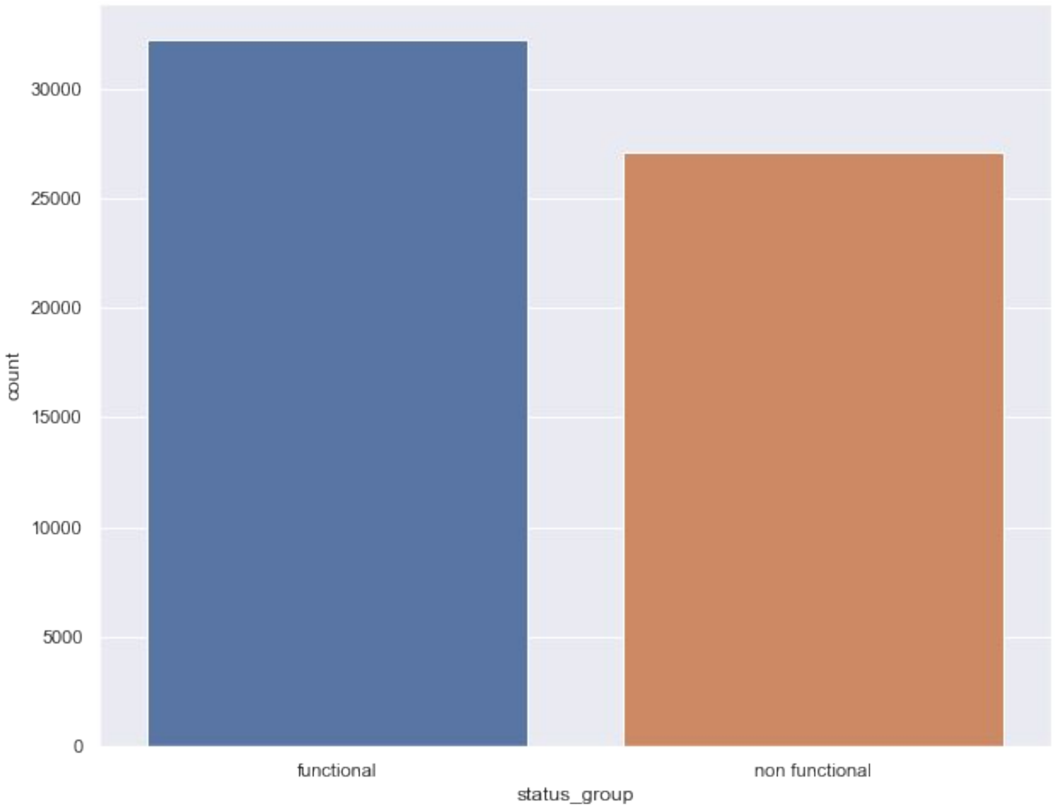
ML Classifier?

- Information gathered between 2011 and 2013 by the Ministry of Water was made available from DrivenData and a partnership with Taarifa as part of a machine learning competition

- Map generated using QGIS software shows the location of the pumps. A green dot indicates a functional pump while a red dot indicates a non-functional pump. Rivers added using USGS shapefile.



Binary Classification Problem. 54% of the 60k wells are functional and 46% are non functional.

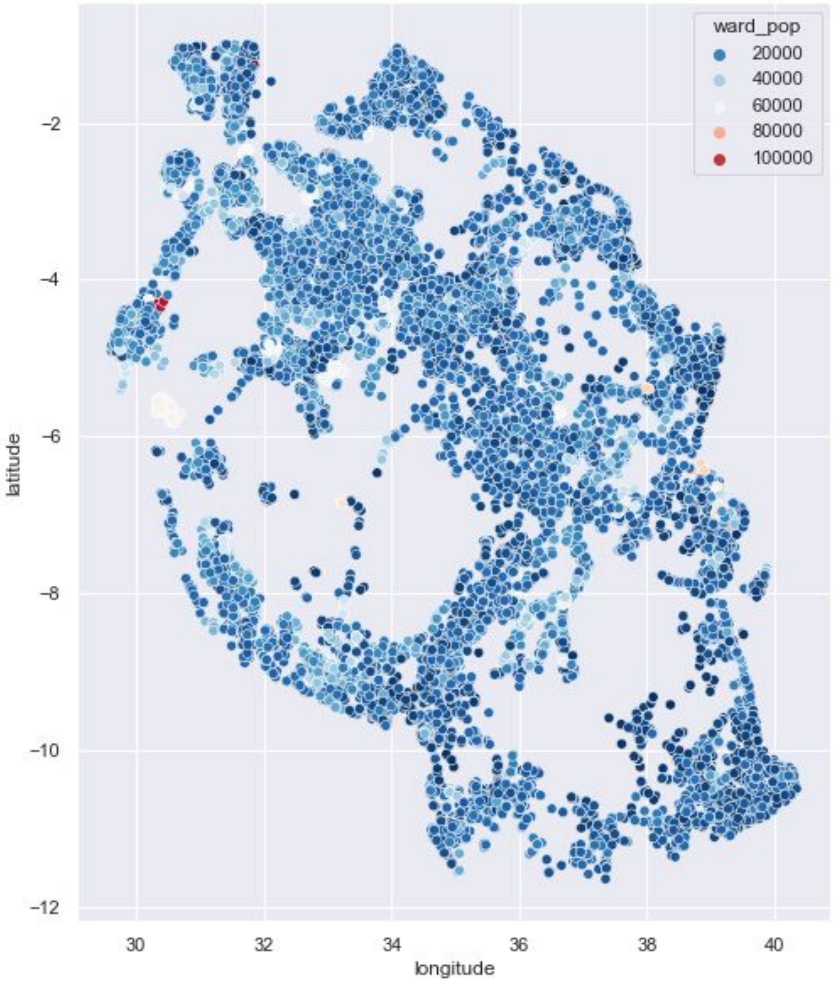


- Population data collected by government census in 2012 was added to the dataset

- ‘Ward’ is a high level geographical division. Thousands of wards in Tanzania

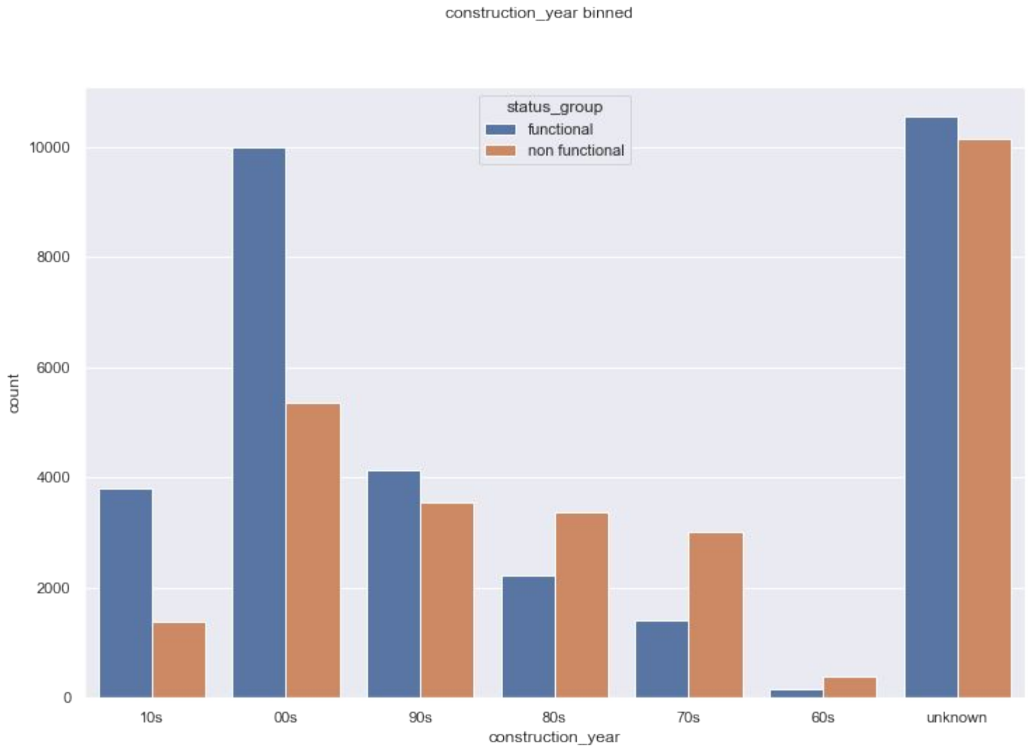
- Population data given in dataset from DrivenData was spotty at best

- Useful for understanding how many individuals could benefit from efficient upkeep of water pumps



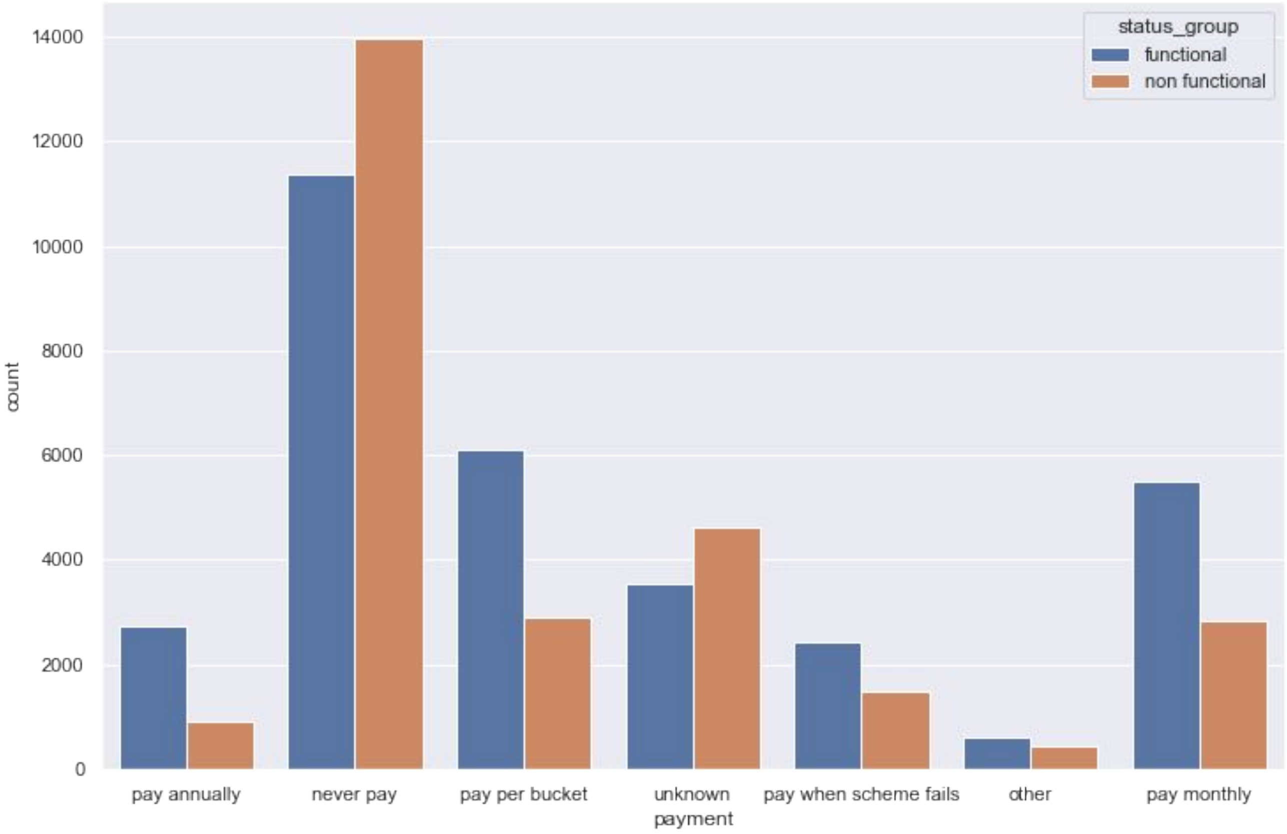
- Older pumps more more likely to be non-functional

    - Water pumps constructed in the 80s or later more likely to be non-functional than functional



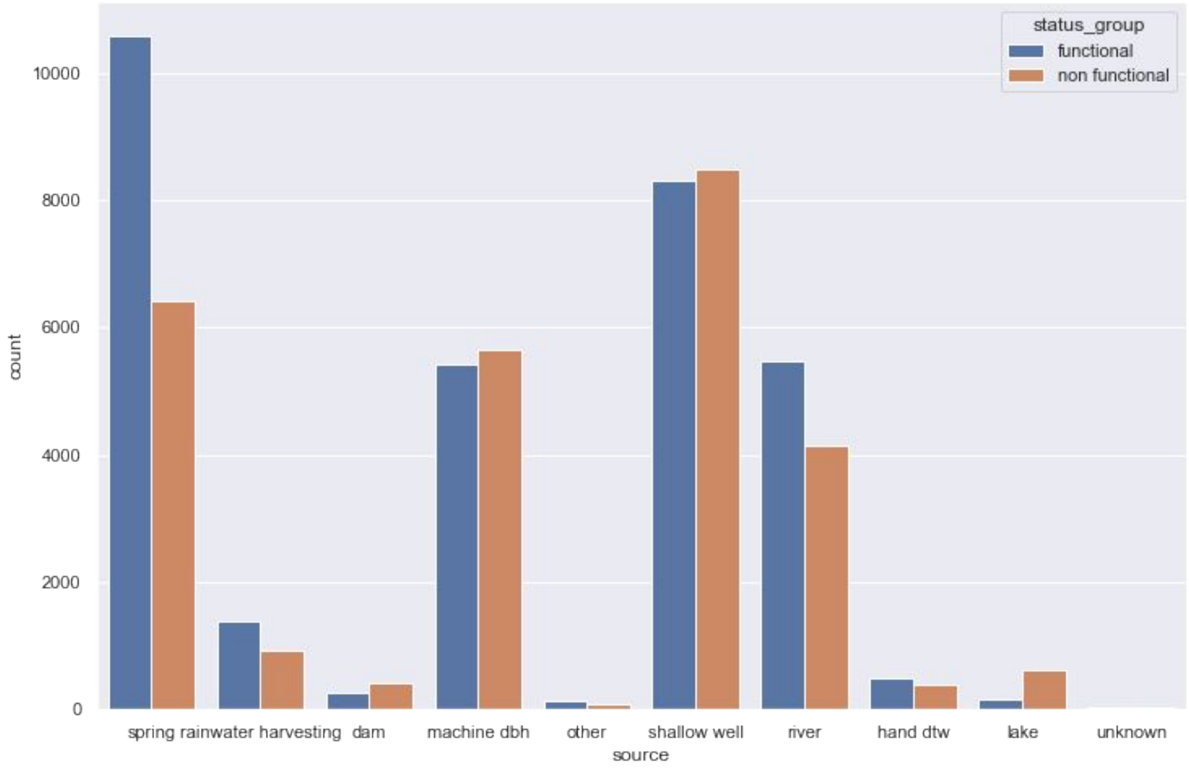
payment information

- proper funding to well critical to maintain functionality

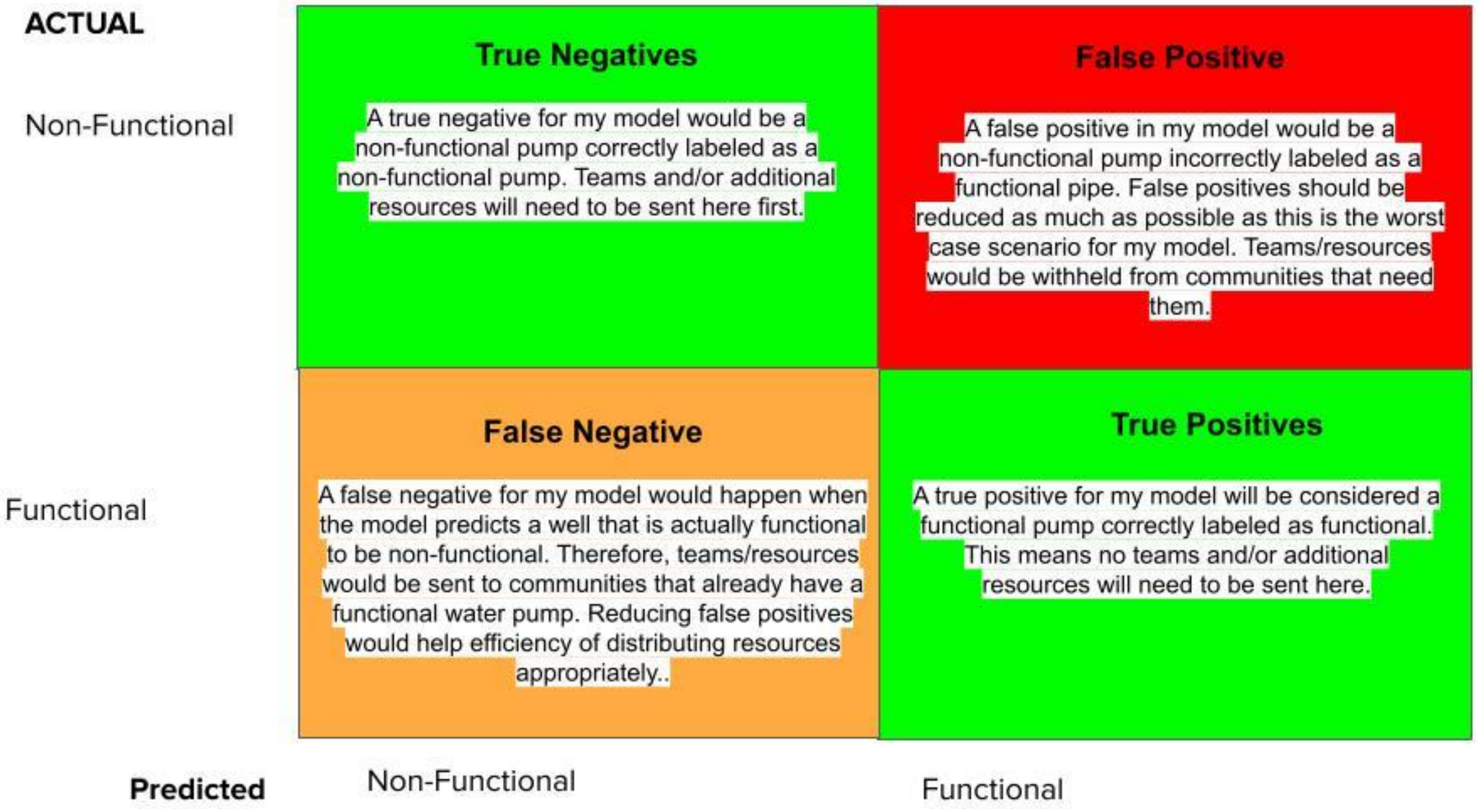


water source

- Source of water for the pump is another critical factor



Model Considerations



ACTUAL		
Non-Functional	**True Negatives** A true negative for my model would be a non-functional pump correctly labeled as a non-functional pump. Teams and/or additional resources will need to be sent here first.	**False Positive** A false positive in my model would be a non-functional pump incorrectly labeled as a functional pipe. False positives should be reduced as much as possible as this is the worst case scenario for my model. Teams/resources would be withheld from communities that need them.
Functional	**False Negative** A false negative for my model would happen when the model predicts a well that is actually functional to be non-functional. Therefore, teams/resources would be sent to communities that already have a functional water pump. Reducing false positives would help efficiency of distributing resources appropriately..	**True Positives** A true positive for my model will be considered a functional pump correctly labeled as functional. This means no teams and/or additional resources will need to be sent here.
Predicted	Non-Functional	Functional

Your services

Increase f1-score

Decrease number of false positives

Decrease number of false negatives

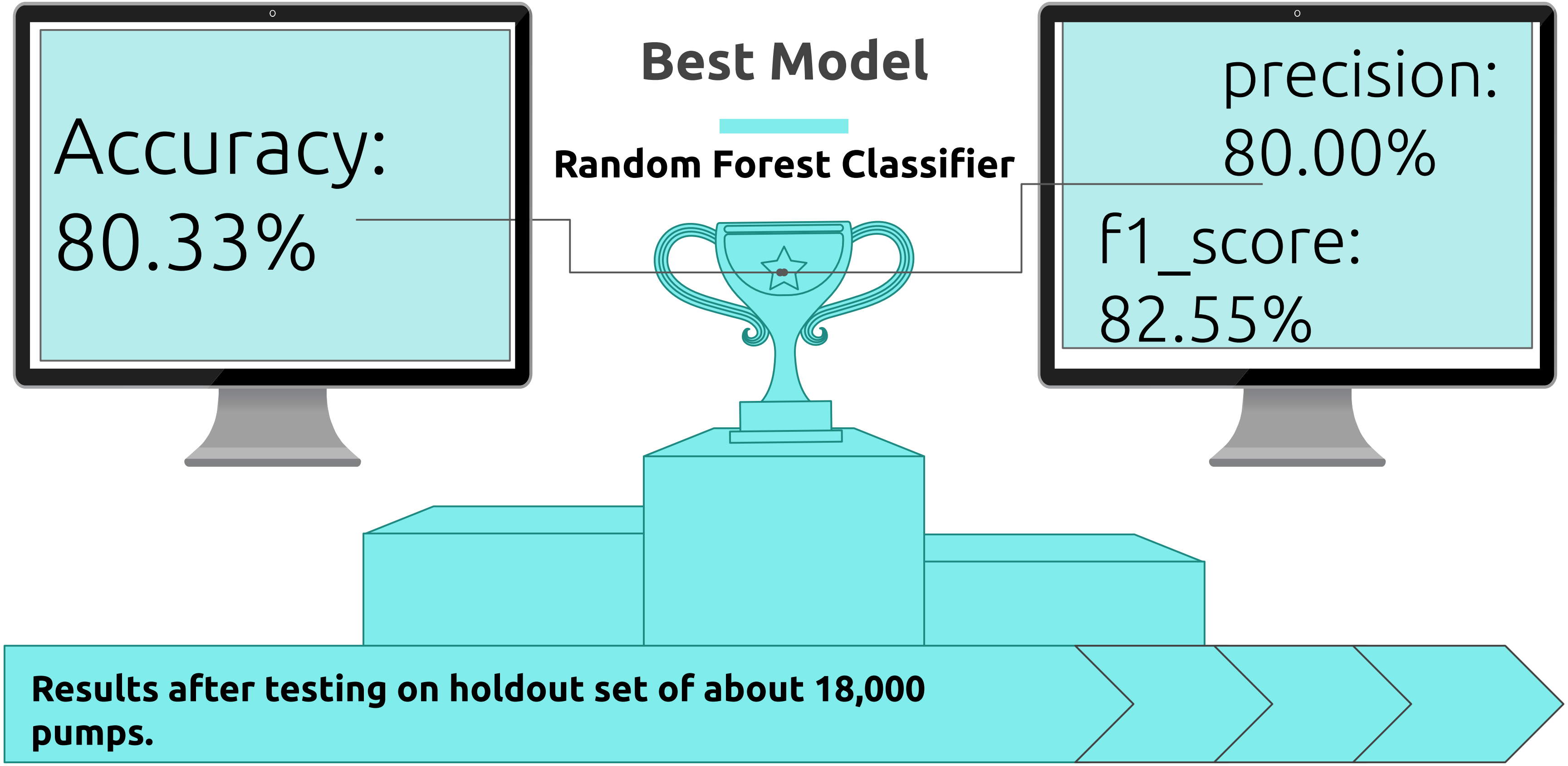
Increase accuracy

Increase number of true positives

increase number of true negatives

Best Model

Random Forest Classifier

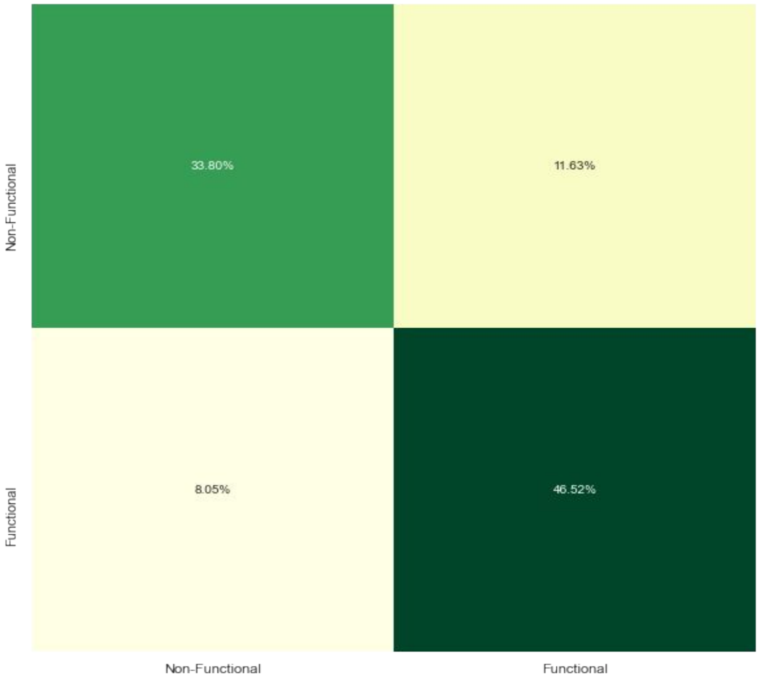


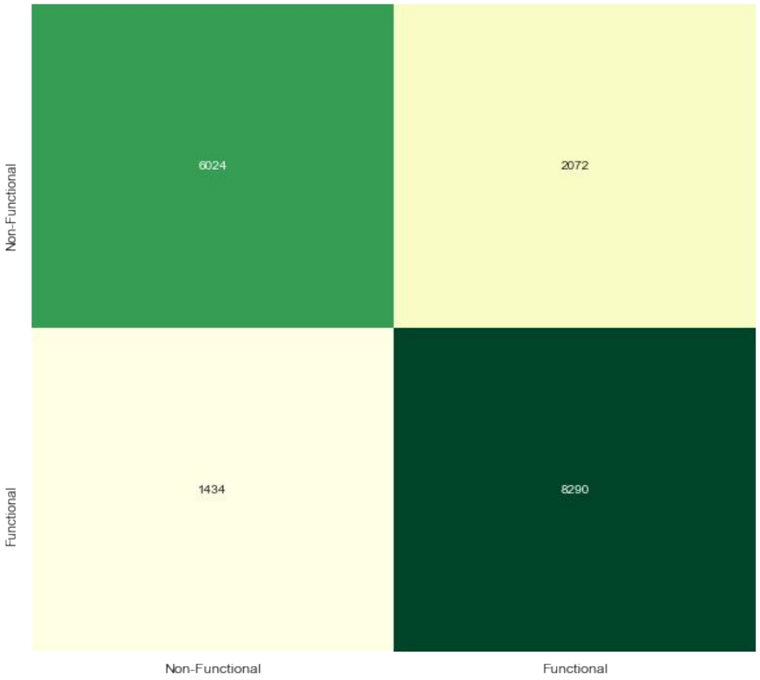
Accuracy:
80.33%

precision:
80.00%

f1_score:
82.55%

Results after testing on holdout set of about 18,000 pumps.





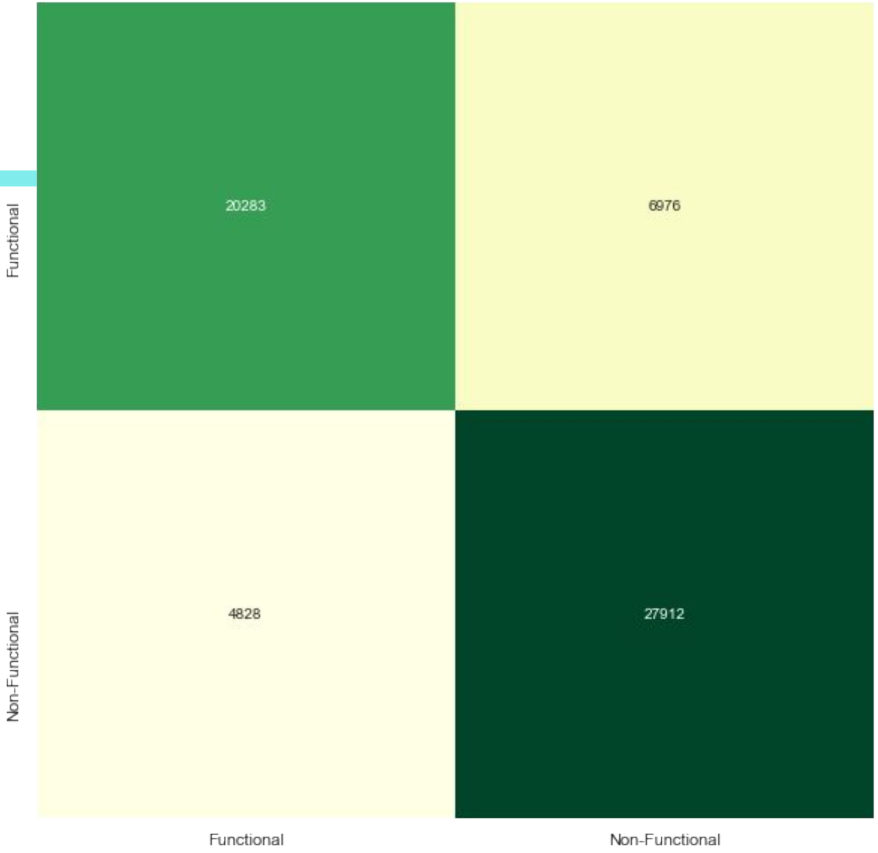
The model predicts a total of 7458 pumps out of the 17820 pumps need repair. It was incorrect only 8.05% of the time when predicted for non-functional pumps. The model also predicts that 10362 of the pumps are functional. It was incorrect 11.63% of the time when predicted for functional pumps.

- A confusion Matrix for my Random Forest Classifier if it were to make predictions on 60,000 wells (the number of wells it took 3 years to visit and record the functionality of by the Ministry of Water).

- It correctly identifies the same amount of pumps in one click as functional as the Ministry of Water did at max capacity in 2011.

- 1164% of all wells unfortunately missed and won't be fixed using the model

- Call to action for ~25k pumps with only a 8.05% rate of deployment to functional pumps instead of non-functional (inefficiency of resource deployment)



Conclusions

- Using the Random Forest Classifier to perform predictive maintenance of water pumps in Tanzania should help the Ministry of Water with resource management.

- The amount of resources saved from correctly identifying pumps with an 80% accuracy rate should more than make up for the resources lost through misclassification of functional pumps as nonfunctional by the model (only 8.05% misclassified in this way).

- Misclassifying 11.63% of non-functional pumps as functional is not ideal. However, the reality is resources are finite. The amount of resources saved from using machine learning instead of physically checking all the pumps should lead to more communities having access to water than before the implementation of predictive maintenance.

Future Work

Use more sophisticated methods to deal with class imbalance and create a tertiary model.

Try more hyperparameter tuning for an XGSboost classifier.

Explore using more geographical and economic data as predictors

Thanks!

Questions?

ddey2985@gmail.com
210-885-7314