TANZANIA MINISTRY OF WATER

PREDICTING THE CONDITION OF WATER PUMPS IN TANZANIA

A Presentation by Patrick Anastasio

We believe in a world where every single person has access to clean and safe water

Problem

- Tanzania is a developing country with varying geography and access to natural resources
- Water is a scarce resource
- Over 59,000 pumps have been installed
- Maintenance is regularly required

Tanzania Ministry of Water

Task

Create a model to predict the condition of a well; Functioning or Non-functioning

- Strategically locate repair teams
- Mobilize quickly when resources and repairs are needed
- Supply needed potable water to villages whose pump is non-functional until repairs can be done

Data Understanding

The data set is provided by Taarifa which aggregates data from the Tanzania Ministry of Water on the over 59,000 water pumps

Method:

- Determine relevant features for prediction
- Create several models and look for best predictor metrics
- Use validation methods to ensure performance
- Tune final model for optimal predictions

Tanzania Ministry of Water

Model Selection

False Positive -
predicted as non-functional but is functional

False Negative -
predicted as functional but is non-functional

Maximize Recall / TPR

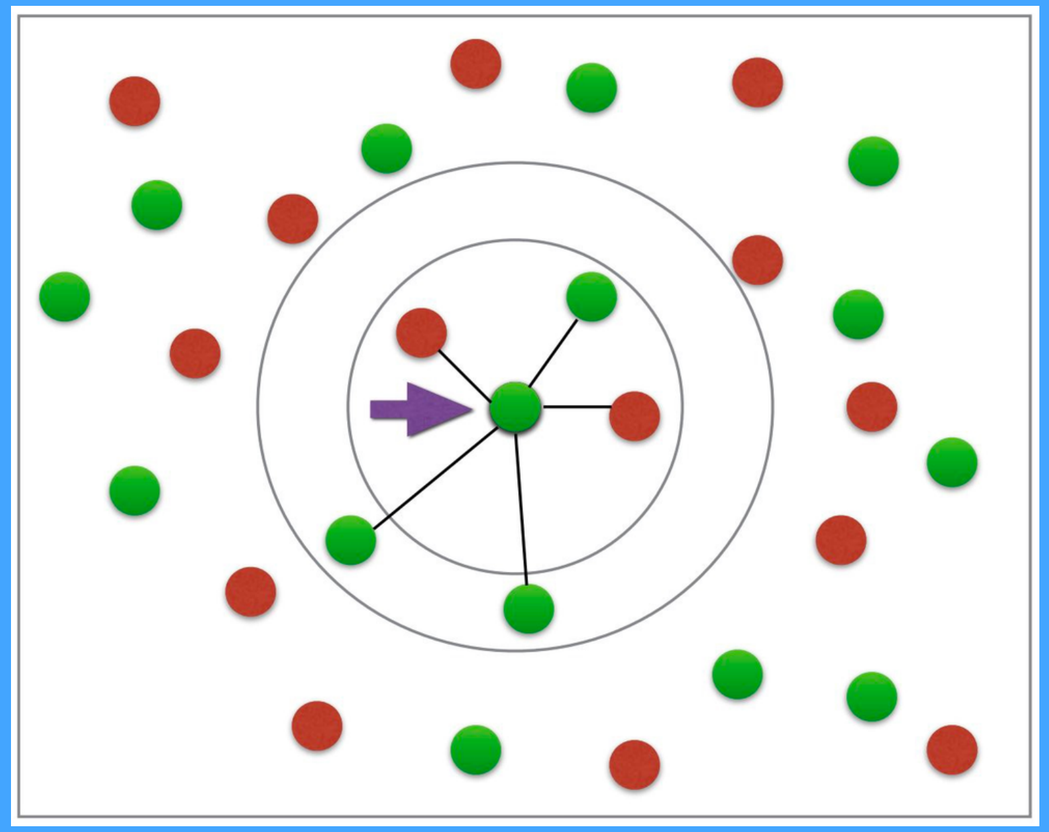
- Minimize false negatives
- Resources are directed to the people who need them
- More human lives are saved

Maximize Precision

- Minimize false positives
- Resources are not directed where they are not needed
- Less logistic strain on the system
- Resources not spread too thin

Inverse Relationship

K NEAREST NEIGHBOR



Most common class among a datapoint's "K" number of nearest neighbors

"K" must be odd

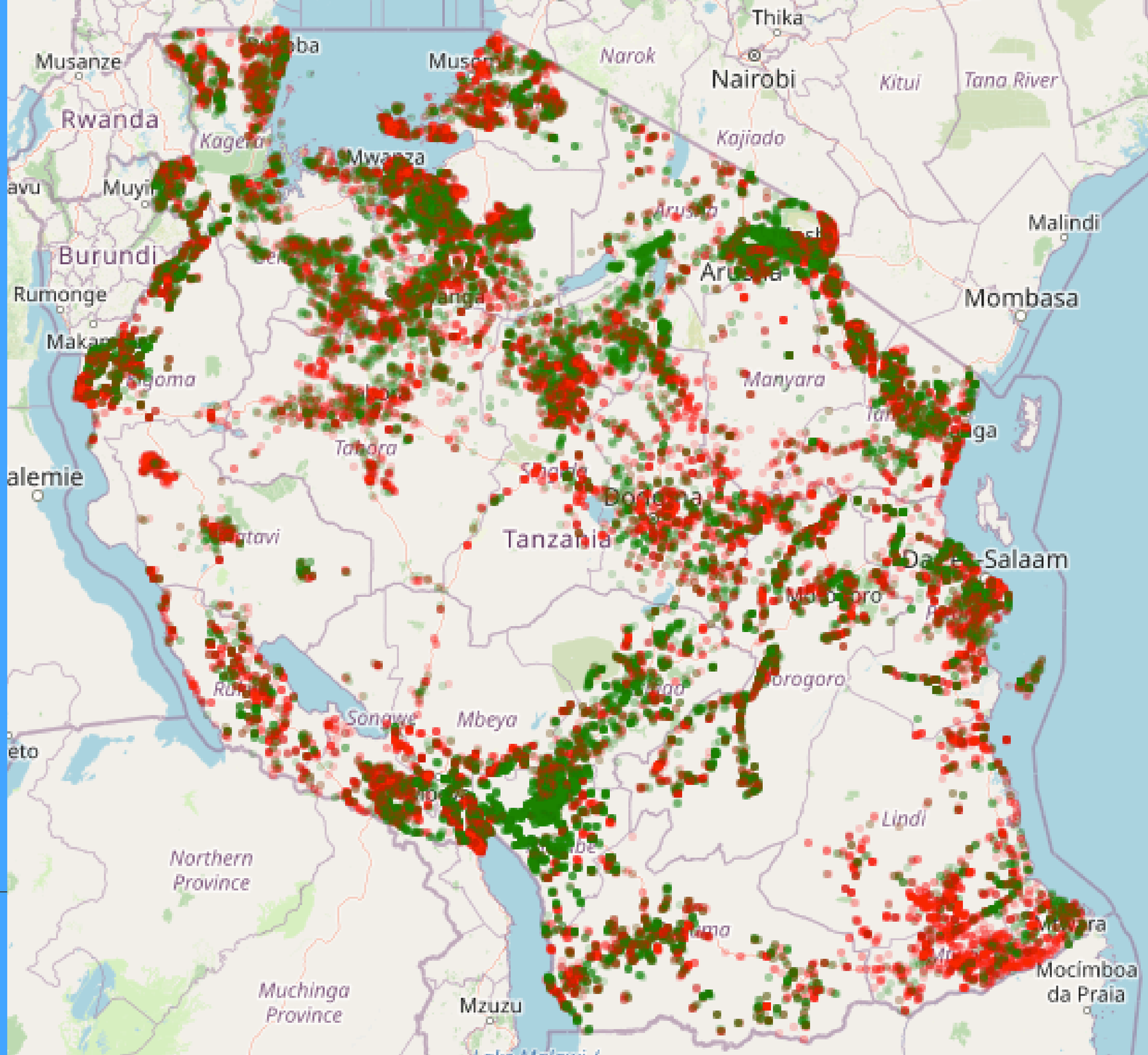
Default is 5

Current Pump Condition

Functional

Non-functional

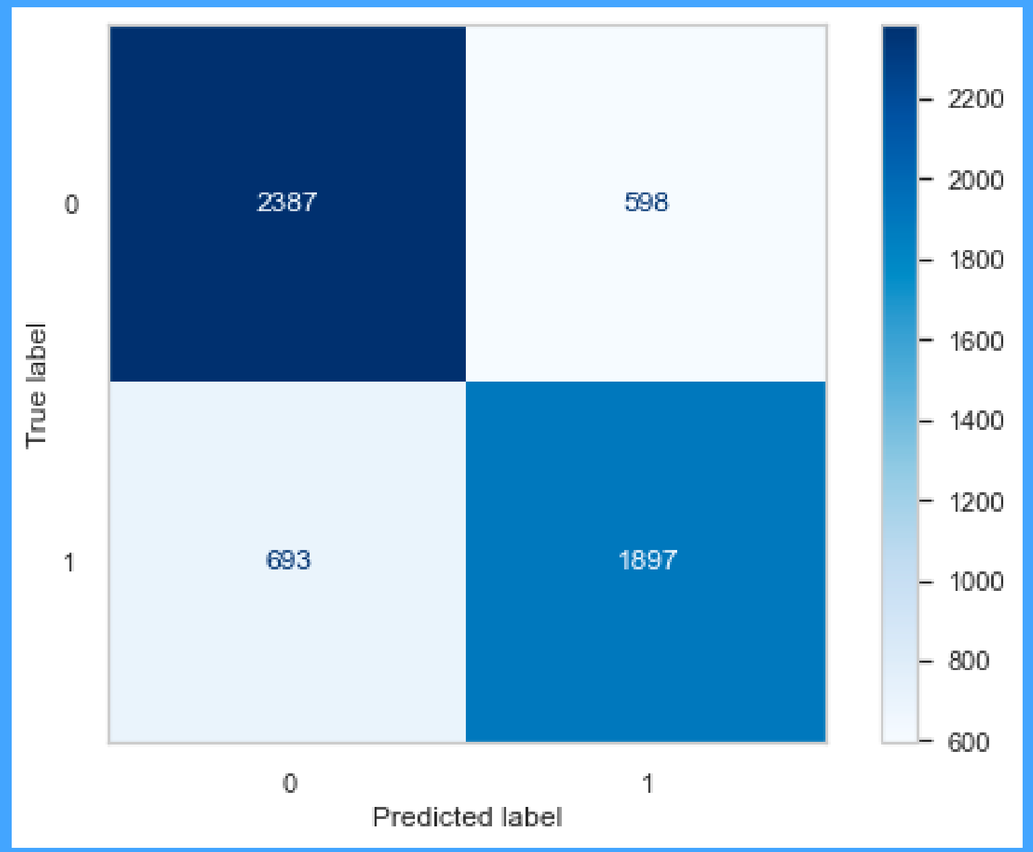
Tanzania Ministry of Water



FINAL MODEL

K Nearest Neighbor

Best Parameters*: (n_neighbors = 5, metric = 'distance')



0 - Functional 1 - Non-functional

- Recall : 0.7324 (+1.89%)
- Precision: 0.7603 (-0.74%)

Actual Non-functional wells are being accurately predicted at a rate of 73.24%

*TUNING RESULTS WERE NOT SENSITIVE TO LEAF_SIZE

RECCOMENDATIONS

- Track seasonal droughts, water conditions and basin levels
- Know what non-functional pumps are un-repairable to better target efforts
- Install more pumps
- Gather more data

NEXT STEPS

- Gather more data
- Re-assess features for significance
- Further tune the model
- Assess more advanced and resource intensive models

Q & A

PATRICK ANASTASIO

- SUDOMAKECOFFEE1@GMAIL.COM
- LINKEDIN: /PATRICKANASTASIO
- GITHUB: PATRICK-ANASTASIO
- MEDIUM: @PATRICKANASTASIO