Water Enhancement Project

Aqua Partners

Stakeholders & Objective

Main Stakeholder: Mr. Gerson Lwenge, Minister of Water and Irrigation, Government of Tanzania

Problematic: Improving the water pump system in Tanzania by identifying the optimal features of installment

Approach: Utilisation of classification algorithms to predict if new water pumps will be functional

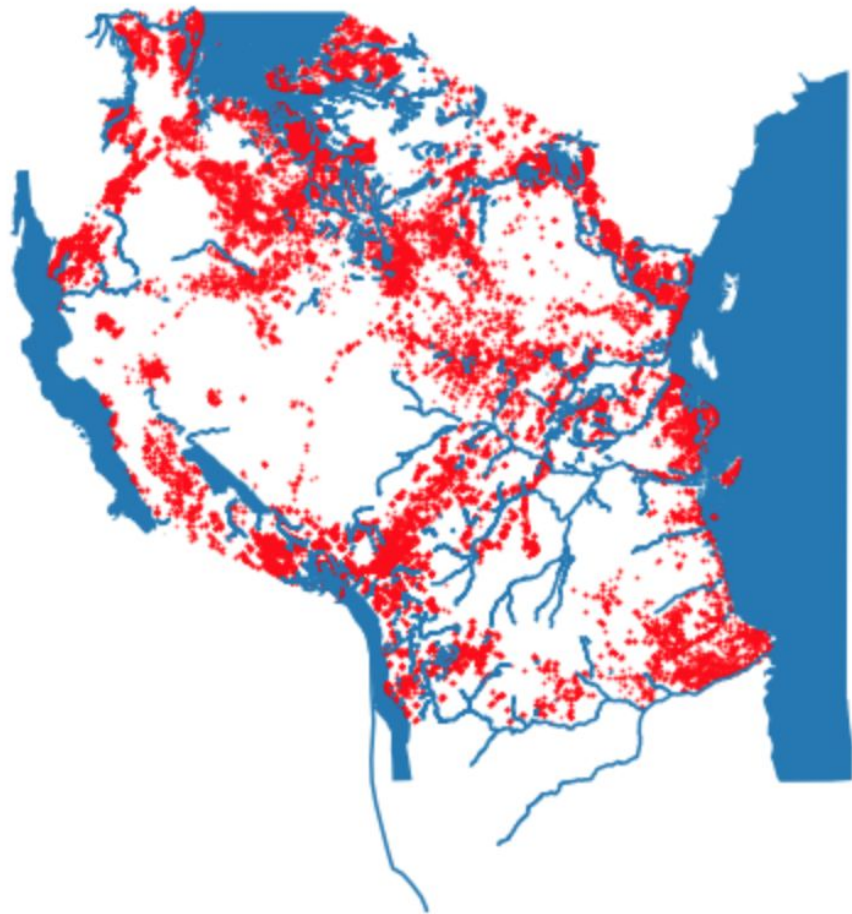# Premise

58
million
people

ECONOMY

6-7% of GDP growth, one of the fastest growing in Africa. Water access is key to sustain the growth and improve people's lives

WATER ACCESS

25 million people lack access to a safe water source and 40 million lack access to improved sanitation

Wells/Water Pumps

Data Details

Approx. 60,000 water pumps installed

Analysis:

- functional/non-functional assessment
- 29 binary & continuous features

Example:

- Date of installation
- Water quality
- Region
- Installing
- Organisation
- Source type
- Nearest water basin
- Length of operation (featured engineered)

Model Process

- <u>Measure Metric:</u> Accuracy

- <u>Rationale:</u>
  - Accuracy will enable us to assess the quality of our model for predicting the status of water pumps

- <u>Validation format:</u>
  - Five-fold Stratified random sampling

- <u>Initial Model Eval:</u>
  - *Best performing:* Decision Tree adjusted using GridSearchCV
  - *Most constant across train/validation:* Logistic Regression

Model Interpretation & Results

- **Baseline Model (train/test):**
  - Decision Tree: 0.96 / 0.78
  - Logistic Regression: 0.78 / 0.77
- **Secondary Model (train/val/test):**
  - Random Forests: 0.77 / 0.76
  - Support Vector Machines: 0.61 / 0.60
  - K-nearest neighbors: 0.71 / 0.69
  - Ensemble methods:
    - Decision Tree with parameter optimization: 0.90 / 0.86 / 0.88
    - Threshold of 0.45

Feature Importance