Predicting Water Pump Functionality in Tanzania

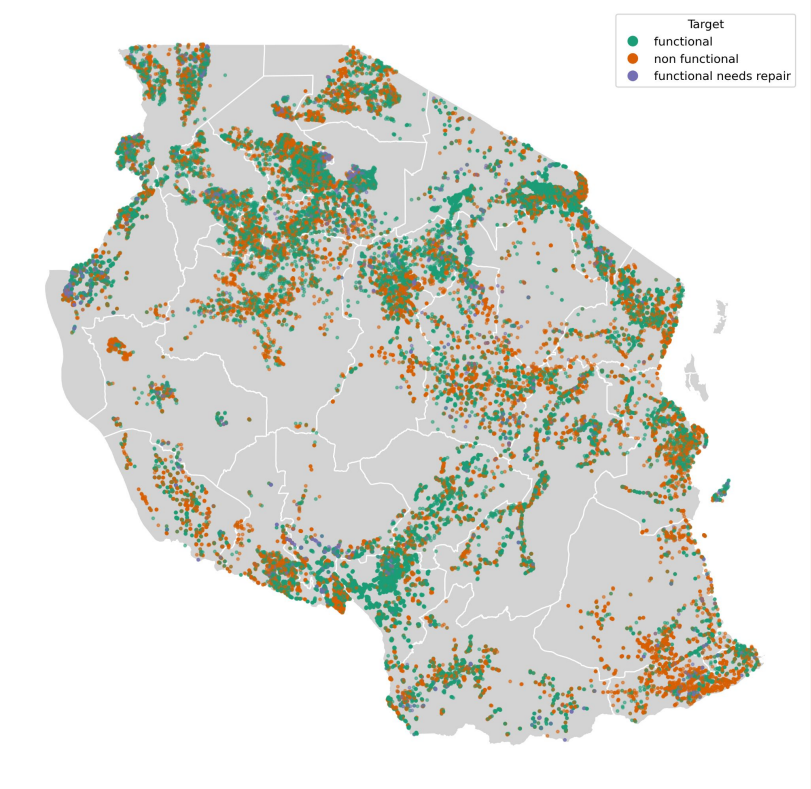
A ML Masterclass by William Guo

Overview

- Build a classification model that can predict whether a water pump is functional or non-functional based on various features
- Data from Taarifa and the Tanzanian Ministry of Water
- Gain a smart understanding of which waterpoints will fail which can improve maintenance operations and ensure that clean, drinkable water is available to communities across Tanzania

Data & Methodology

- 39 features → 18 features
- Multi-class target variable → binary target variable
- 5 classification algorithms trained:
  - Logistic Regression
  - Decision Tree
  - Naive Bayes
  - Random Forest
  - XGBoost



Modeling & Evaluation

Models with best accuracy score

- Random Forest: 0.8119
- XGBoost: 0.8097

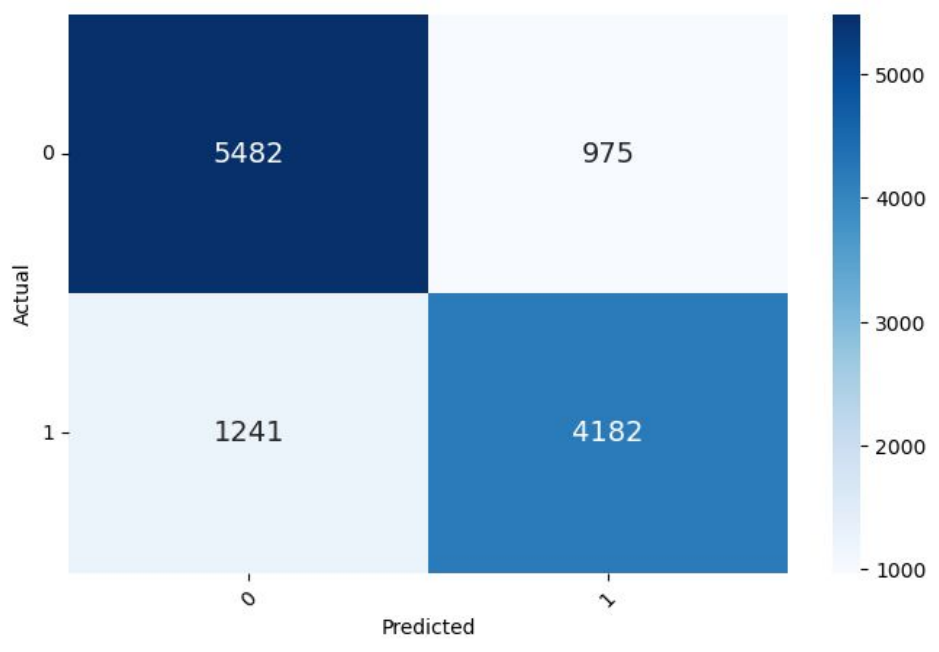
Hyperparameter tuning for RF

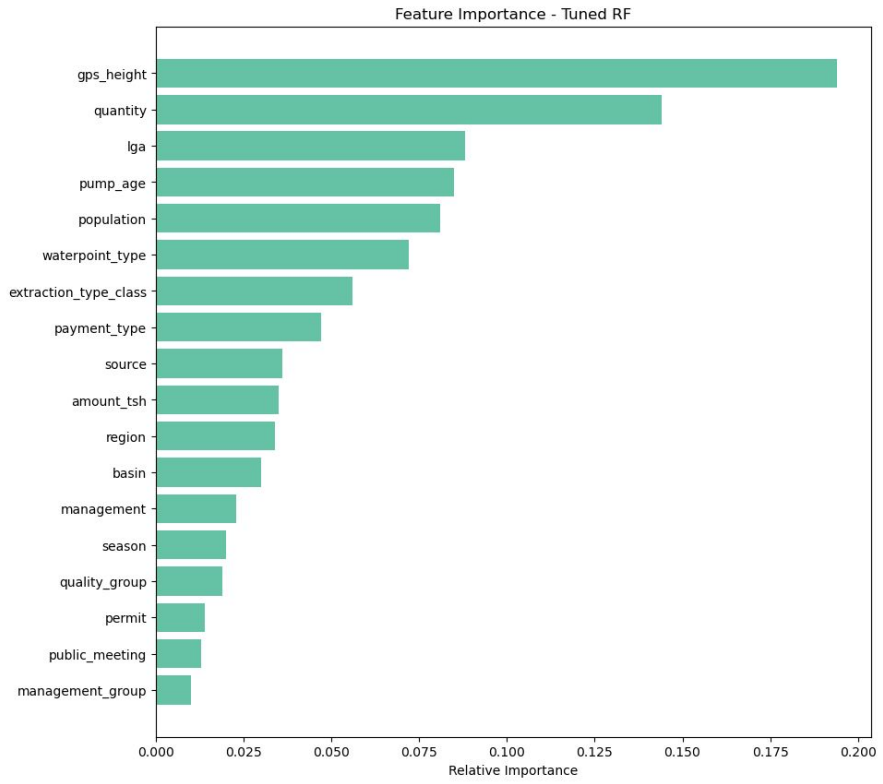
Hyperparameter optimization led to a very slight increase in model accuracy.

81.34%

Tuned RF Classifier Accuracy Score

Tuned RF Classifier Summary





Thanks!

Do you have any questions?

willi.guo@gmail.com
+1 646 298 6341
github.com/williambguo

CREDITS: This presentation template was created by Slidesgo, and includes icons by Flaticon, and infographics & images by Freepik