

A map of the African continent is shown in a dark grey tone. The country of Tanzania, located in East Africa, is highlighted in a dark red color. The title text is overlaid on the map.

# *Water Pump Functionality Prediction in Tanzania*

*BY  
FRIDAH KIMATHI*

# ***OVERVIEW***

The project aims to develop a model to classify the functionality status of water pumps in Tanzania using data sourced by Taarifa and the Ministry of water.



# ***BUSINESS PROBLEM***

Tanzania is facing a water crisis

57 million people struggle to access clean water

The Tanzanian government is trying to resolve crisis by regular maintenance/repair of water pumps

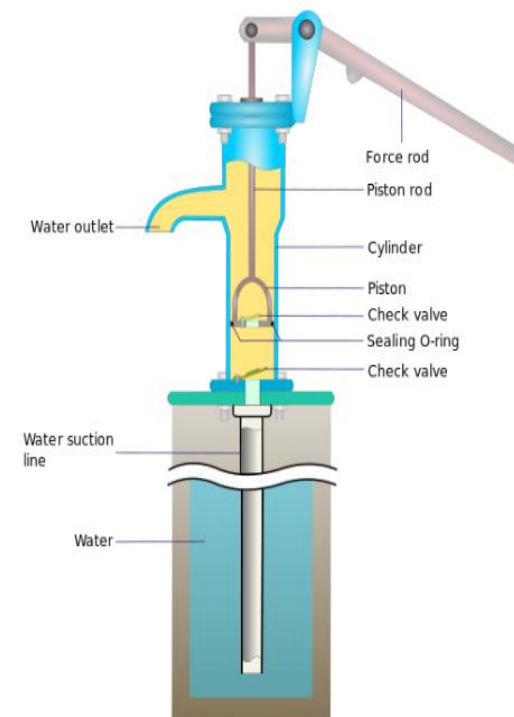
Its engineers are finding it difficult to achieve this objectives due to lack of information on non-functional/failing pumps

# *DATA*

The data used in this project is from the Pump it Up: Data Mining the Water Table competition hosted by DrivenData, originally sourced by Taarifa and the Tanzanian Ministry of Water

## Pump it Up: Data Mining the Water Table

HOSTED BY DRIVENDATA



# ***MODELLING***

|   | <b>Model</b>                         | <b>Train Accuracy Score(%)</b> | <b>Test Accuracy Score(%)</b> |
|---|--------------------------------------|--------------------------------|-------------------------------|
| 0 | Baseline Decision Tree               | 100.0                          | 75.0                          |
| 1 | Second Decision Tree                 | 100.0                          | 75.0                          |
| 2 | Baseline Random Forest Classifier    | 0.0                            | 0.0                           |
| 3 | Baseline Gradient Boost              | 99.0                           | 70.0                          |
| 4 | XGBoost Classifier                   | 86.0                           | 77.0                          |
| 5 | Random Forest Classifier-Grid Search | 98.0                           | 80.0                          |
| 6 | XGBoost Classifier-Grid Search       | 95.0                           | 79.0                          |
| 7 | Final Model-Random Forest Classifier | 98.0                           | 80.0                          |

# *EVALUATION*

- The accuracy score of the model is 0.7937.

**Upload new submission**

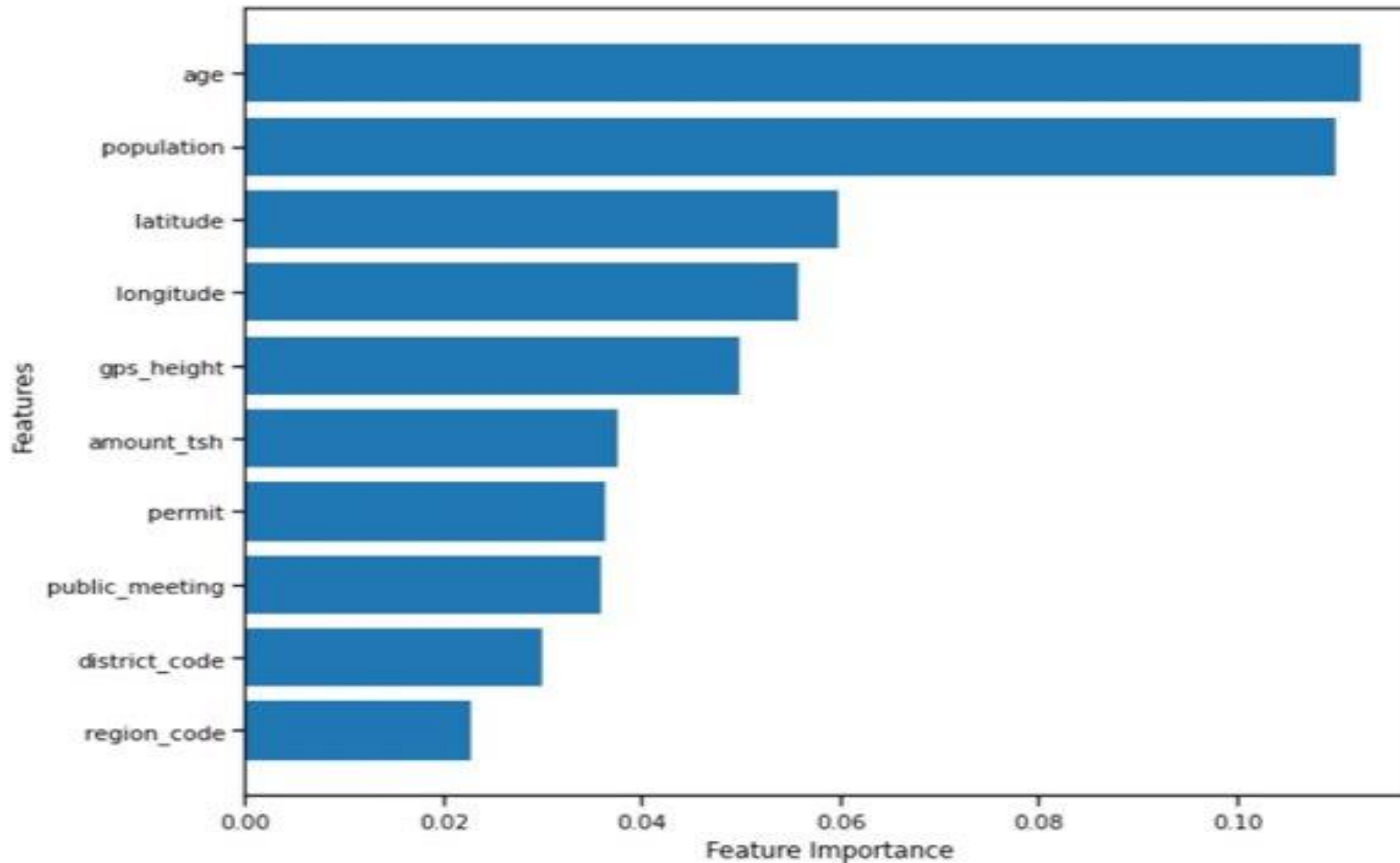
Woohoo! We processed your submission!

Your score for this submission is:

0.7937

# *EVALUATION*

The top 10 most important features are:



# ***LIMITATIONS***

The overall data used was not up to date and was not reliably gathered.



It contained a lot of placeholders in important features such as 'population' and 'amount\_tsh'.



Hence the results obtained are not particularly accurate.





# ***RECOMMENDATION***

Work with the local government to ensure more accurate gathering of data

Data collected should highlight more on non functional pumps or those in need of repairs

Data collected could highlight functional pumps that are unlikely to fail

A close-up photograph of a person's hand filling a yellow plastic water container from a public tap. The tap has a blue handle and is mounted on a wooden structure. The person is wearing a patterned shirt. The background is slightly blurred, showing a wooden surface and some tools. The text "THANK YOU" is overlaid in the center in a white, italicized font, with a small red horizontal line underneath it. There are decorative pink and red dots in the bottom right corner.

*THANK YOU*