A stylized map of Tanzania is centered on a light cream background. The map is filled with various shades of blue and green, representing different geographical features. Numerous small, dark blue dots are scattered across the map, representing water wells. The dots are more densely clustered in the northern and eastern regions. The title text is overlaid on the central part of the map.

TANZANIA WATER WELLS PREDICTIVE ANALYSIS

TROY STEVE

In the bottom right corner, there are several hand-drawn, dark blue lines that resemble the outlines of fingers or a stylized signature.

Business Problem

- DEVELOP A CLASSIFIER TO HELP PREDICT THE STATUS OF WATER WELL PUMPS.

- *Functional status*
- *Functional but need repairs status*
- *Non-functional status*

DATA

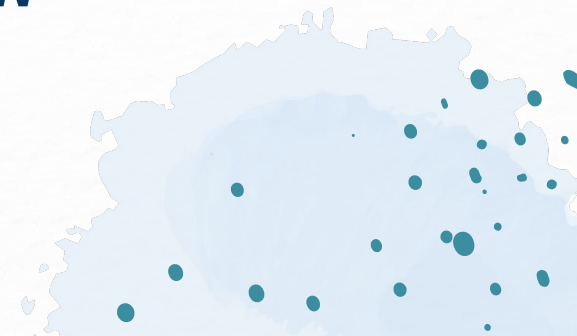
The data at our disposal comes from Taarifa and the Tanzanian Ministry of Water.

The target (`status_group`) has three categories making this a ternary classification problem by default.

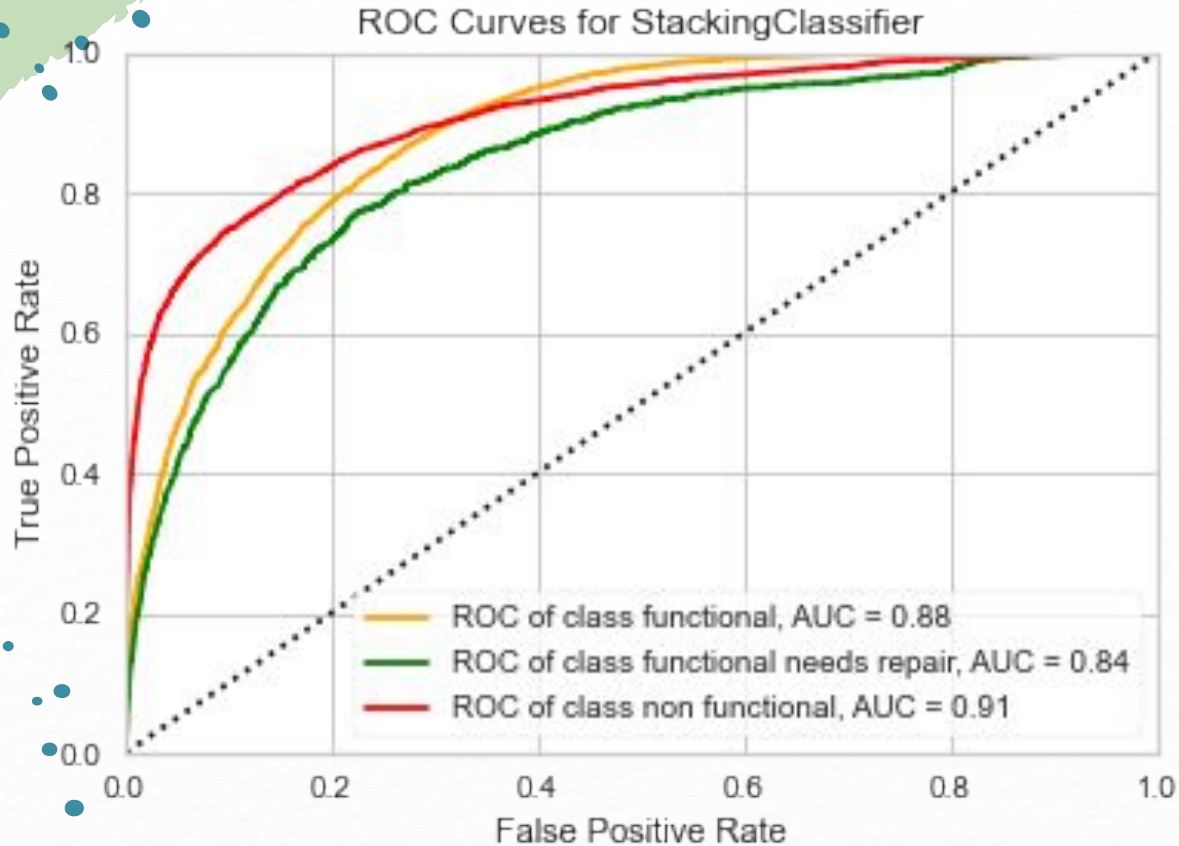


MODEL DEVELOPMENT

- ***DATA CLEANING***
- ***EDA***
- ***MODELING***
 - ***SIMPLE MODELS***
 - ***TUNED AND COMPLEX MODELS***
 - ***MODEL SELECTION***
 - ***INTERPRETATION***

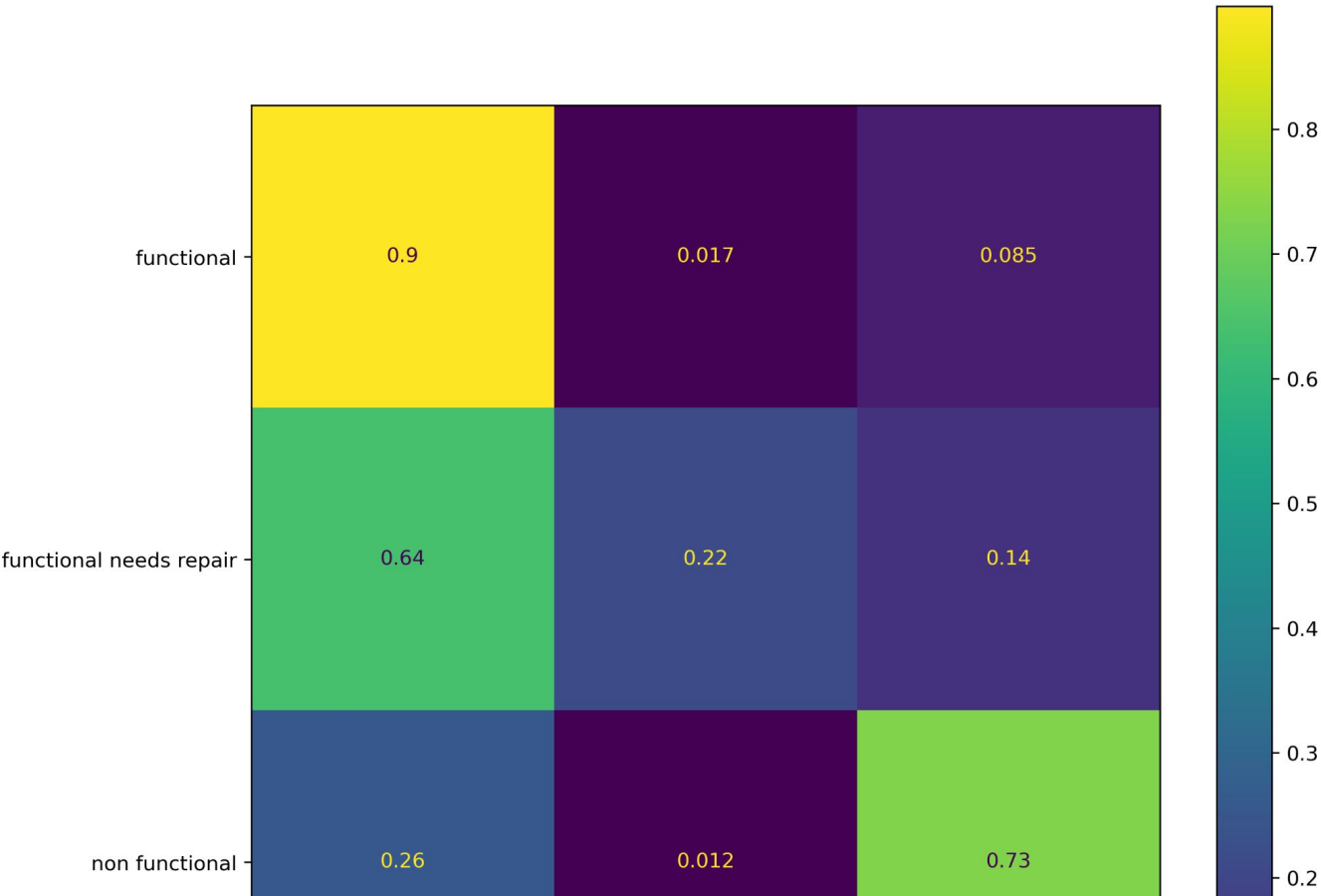


MODEL SELECTION



**THE BEST CLASSIFIER WAS
SELECTED BY CHOOSING THE
ONE THAT HAD THE HIGHEST
F1 SCORE**

Combined XGBoost and Random Forest Classifier



FINAL MODEL

The final model has the following scores

- accuracy: 0.783
- precision: 0.777
- recall: 0.783
- f1: 0.771

Conclusion

- The model we've come up with has an accuracy score of 78% meaning that the Tanzanian government will be able to predict correctly 78% of the times it wants to pump funds towards replacement of pumps that aren't functional and towards repairing of those that are functional but need repairs.
- This will save the government the time and resources it would have used to go round checking the pumps and at times investing in those wells that it shouldn't.