

TANZANIA-WATER-POINTS FUNCTIONALITY PREDICTION USING CLASSIFICATION MODELS

BY: PETER WAITI

Business Problem:

An NGO focused on locating wells that needs repair has been contacted by the Tanzanian Government. The Tanzanian government is looking to find patterns in non-functional wells to influence how new wells are built. As an employee of the NGO, I have been tasked with presenting findings on how the ministry of water can identify wells that are non-functional or needs repair, so as they can dispatch teams to fix them.

About Study Area:

- Tanzania is an East African country with about 59,678,000 in population. It is located just south of the Equator. Tanzania is a developing country that still has struggles with providing clean water to its entire population. There are many water points already established in the country, but some are in great need of repair while others have failed altogether.

Data used:

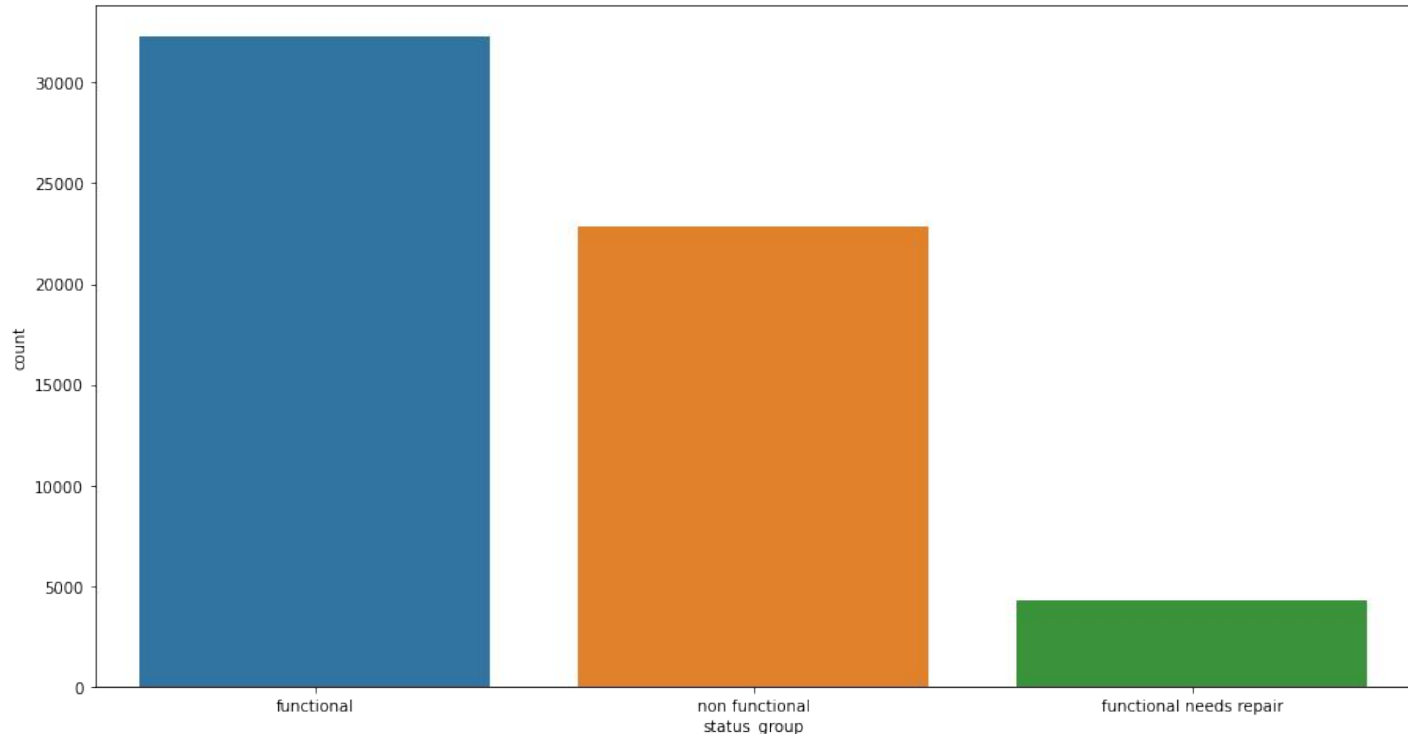
- The data sets contained information about 59400 water points with 40 total features that related to these water points. There are various variables from these data sets that influence the functionality of the water points. Some of the variables explain the location of the water points, functionality of the water points, type of pumps in each well, time they were installed, year the information was recorded e.t.c.

Methodology:

- Understanding the data
- Preparing the data for the model.
- Build different predictive classification models for the functionality of the wells.
- Compare results from each Model.

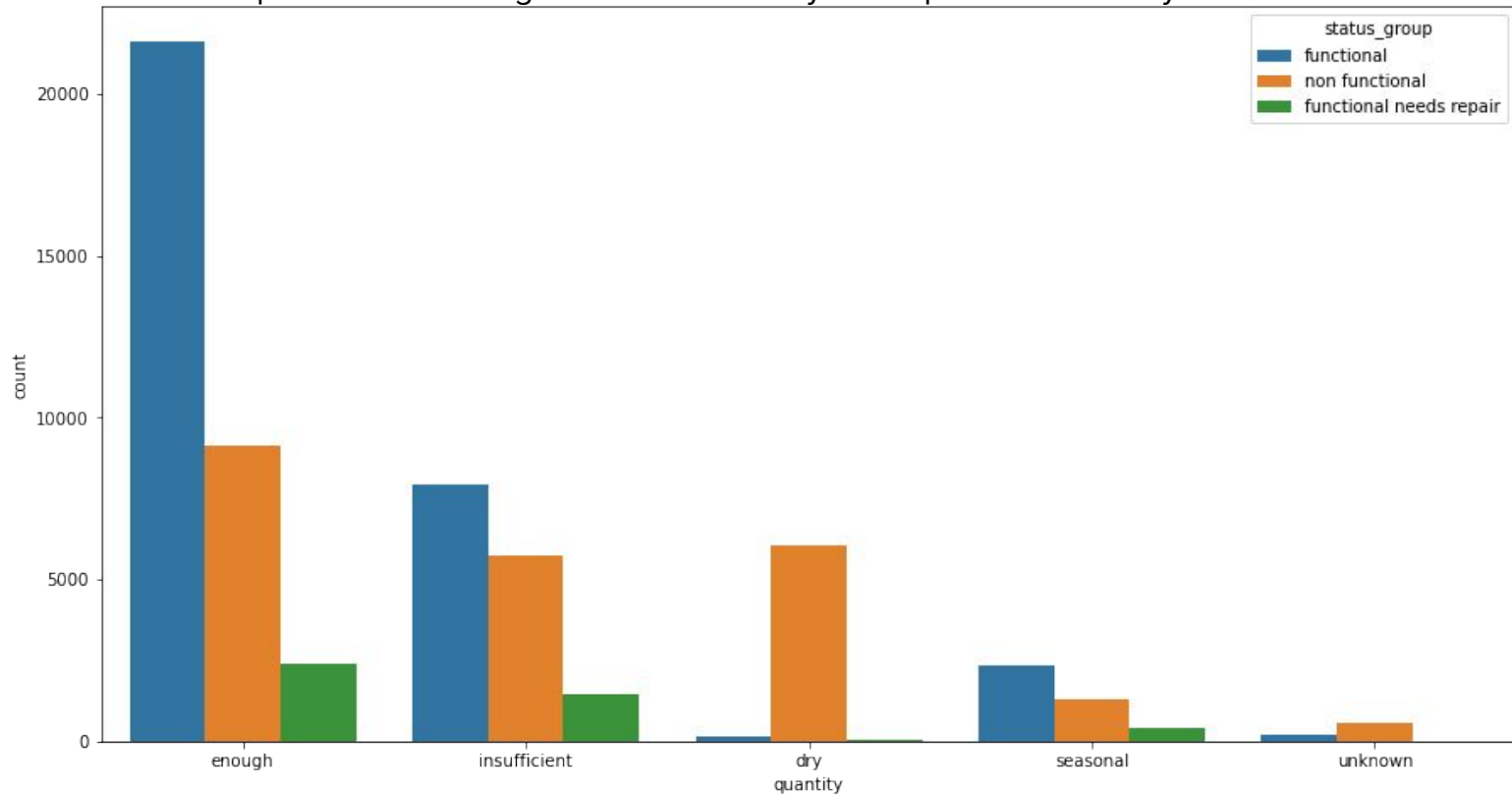
Results:

Distribution of functionality of the wells. From the plot below above 30,000 water points are still functional, about 20,00 are non-functional and about 5000 are functional but needs repair.



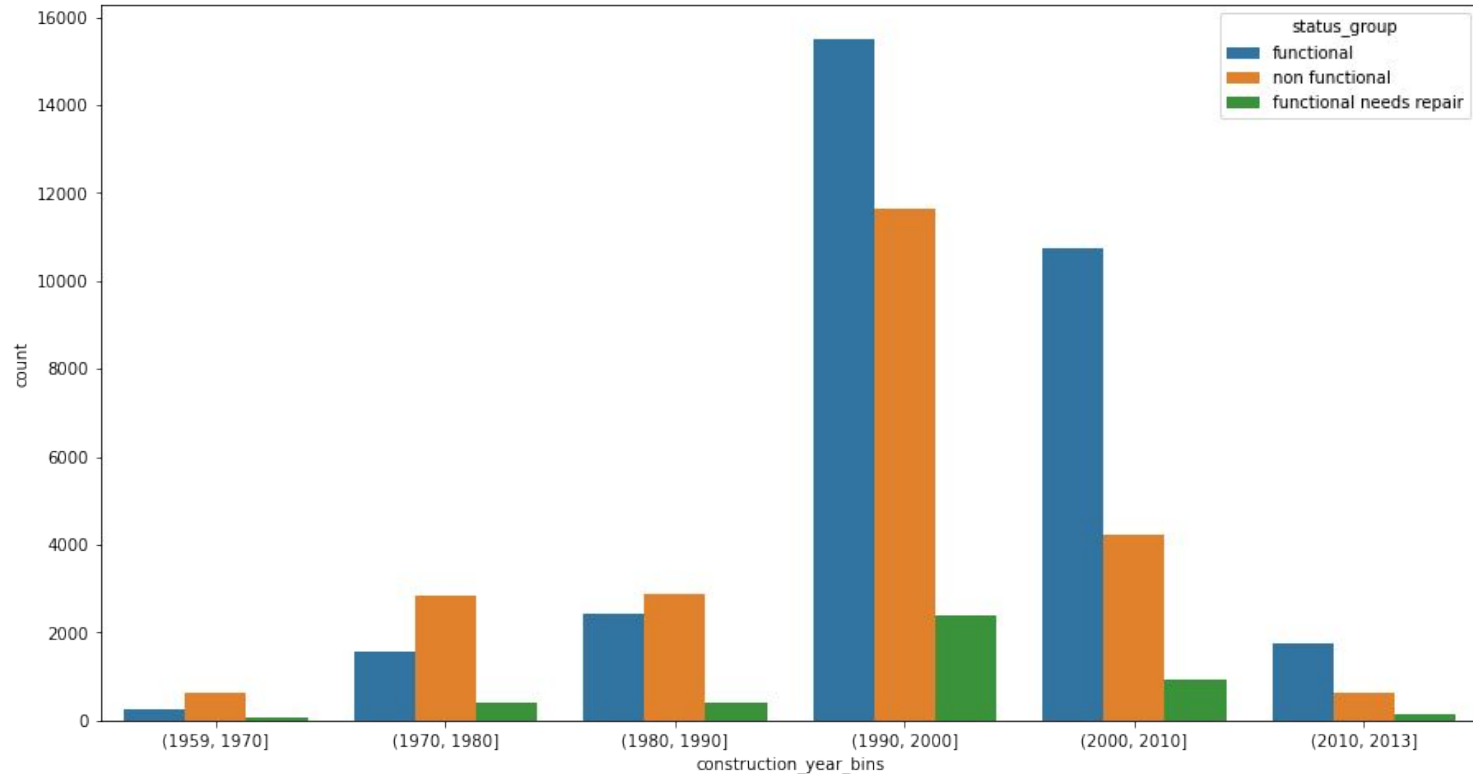
Results:

The graph below shows the quantity of water in relation to functionality of the water points. From the plot you can see that most functional water points have enough water and the dry water points are mostly non-functional.



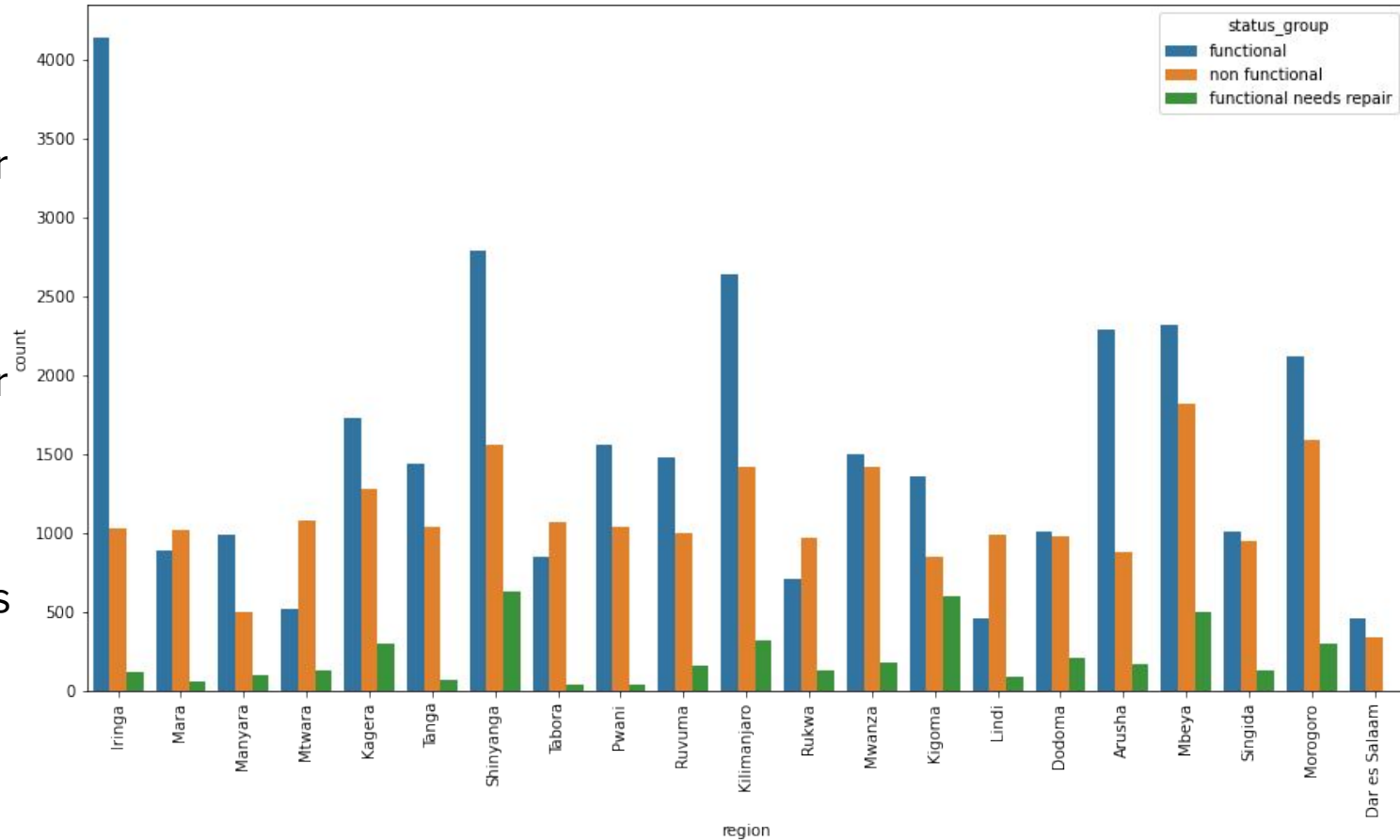
Results:

The graph below shows how age affects the functionality of the water points. From the plot you can see that most recent water points are still functional while the water points constructed before 1990 most are non-functional.

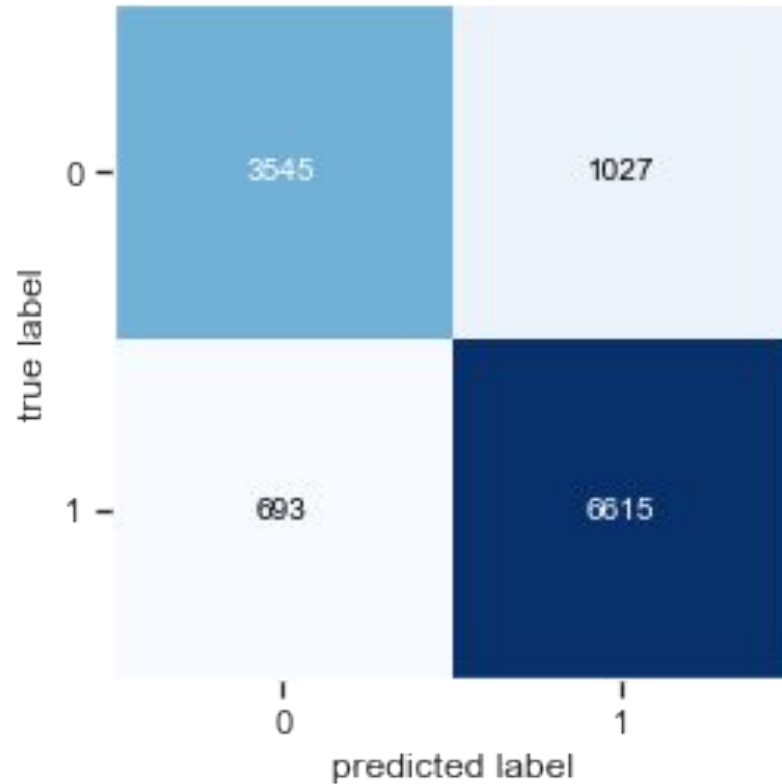


Results:

Iringa region has the most functional water points while dar-es salaam has the least functional water points even though the population of dar-es salaam is high.



Results:



From the Random forest model prediction. From the confusion matrix results 6615 water points were correctly classified as functional while 693 water points were wrongly classified as non-functional but were actually functional water points. Also, 3545 water points were correctly classified as non-functional while 1027 water points were wrongly classified as functional but were actually non-functional.

Findings:

1. Different models were compared and the best mode was found to be Random Forest classifier with an accuracy of 85%.
2. 4272 Wells were dried up but have good water quality.
3. Finding clean water sources is a problem in Tanzania.
4. 2226 wells making up 7% of the total wells have enough and soft water but needs repair. The Tanzanian Government should invest in repairing them otherwise they will be nonfunctional.
5. 8035 wells have enough, good quality water but are nonfunctional. Thus the need to find ways to pump the water from these wells.

Recommendation:

- I would recommend that the Tanzanian Government continue to make recordings of the wells so as to continue monitoring them since the data provided was only recorded mostly 2011 and 2013. This would provide more accurate data for model improvement.
- Different regions have different factors like climate, rainfall seasons etc. so thus different modes should build for each region.

Questions?

Thank You!!