Predicting the Functionality of Water Wells in Tanzania

ZENAH BIWOTT

OVERVIEW

- BUSINESS PROBLEM
- OBJECTIVES
- METHODOLOGY
- DATA INSIGHTS
- FINDINGS
- RECOMMENDATION
- FUTURE IMPROVEMENTS
- CONTACT INFO

BUSINESS PROBLEM

Investment has been done heavily on drilling water wells, but many of these wells are non-functional, leaving millions of people in Tanzania lack reliable access to safe drinking water.

The lack of functional water wells has severe consequences, including:

- increased risk of waterborne diseases
- reduced economic productivity
- decreased quality of life

The goal of this project is to develop a predictive model that can identify which water wells are likely to be non-functional, allowing the Tanzanian government and other stakeholders to target their efforts more effectively, improving the lives of thousands of people.

Our model can be used to:

- inform data-driven decision making
- optimize maintenance and repair efforts
- reduce the number of communities without access to safe drinking water

OBJECTIVES

Develop a predictive model that can accurately classify wells as functional or non-functional

Evaluate the performance of different machine learning models and select the best one

Provide recommendations for improving the accuracy of well functionality predictions.

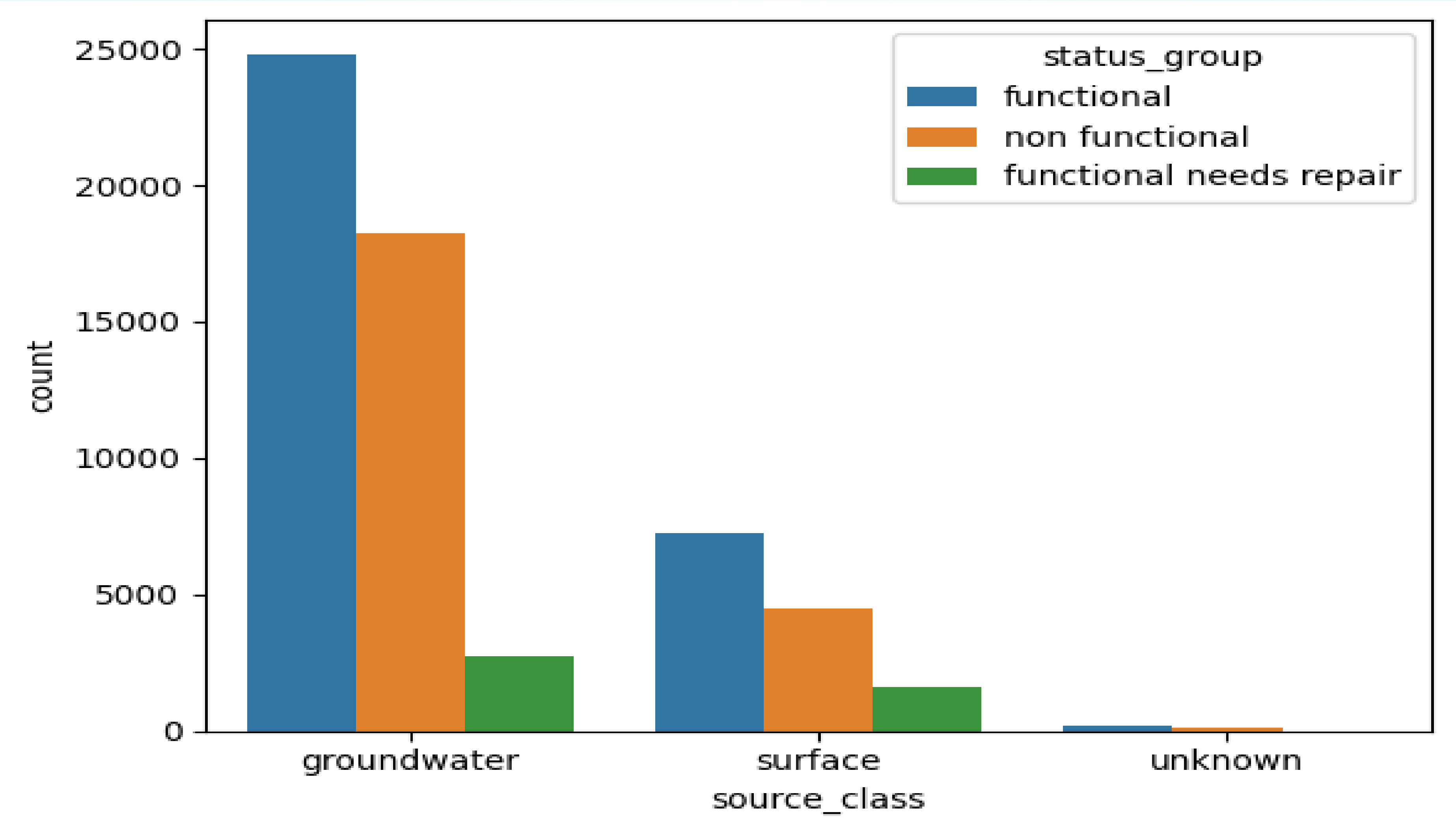
METHODOLOGY

I utilized a supervised classification machine learning algorithm to classify the functional status of water well pumps. Our final model is a decision tree.

DATA INSIGHTS

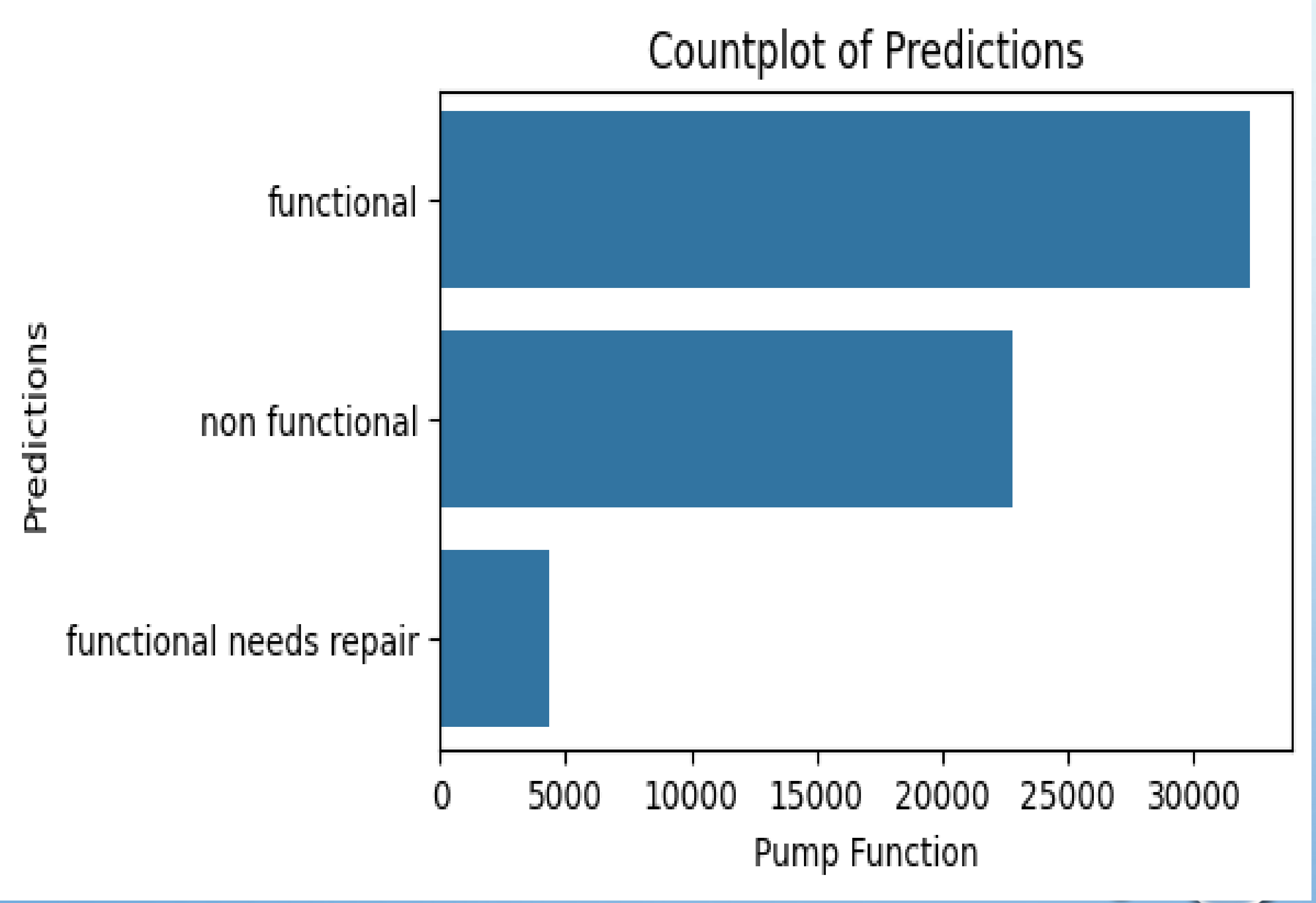
- The data for this project comes from the Taarifa waterpoints dashboard, which aggregates data from the Tanzania ministry of water.
- It is divided into three csv files: a training set containing 59,400 observations (80%) and a test set containing 14,850 observations (20%). The third dataset contains training set labels that detail status group information for each of the training set values indicating whether the pump is “functional”, “non-functional”, or “in need of repairs”. Both the train and test datasets have 40 similar columns with information about water pumps in Tanzania.

Water source

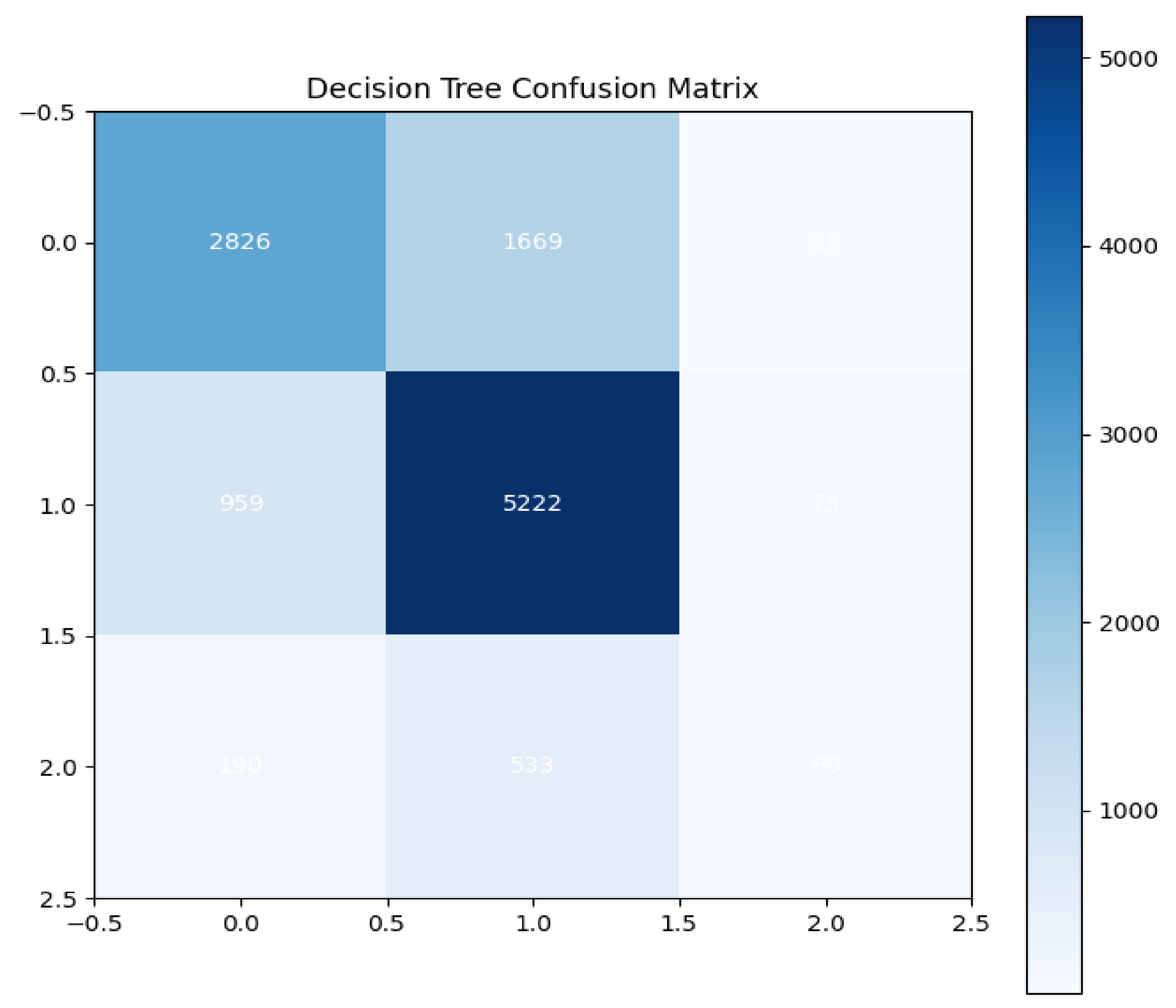


When we look at the columns, there are lots of non-functional ground water

- Distribution of the target variable



- We notice that the target classes are imbalanced, with the functional pumps being the most common at more than 30,000 followed by non-functional pumps at over 20,000 pumps, and the lastly the functional-need-repair pumps at less 5,000.



'non functional': 0
'functional': 1
'functional needs repair' : 2

The confusion matrix shows that the model has a bias towards predicting that a well is functional since it is the most frequent observation, showing 5222 wells being functional.

From our classification report, we can see that the accuracy of our model is 70%. This is a better result that the first model that we used, however, it is still less than our desired accuracy of 75%

FINDINGS

K-nearest Neighbors: The model has an accuracy of 68%.

Logistic Regression: The model has an accuracy of 68%.

Decision Tree: The model has an accuracy of 70%.

The Decision Tree model achieved a is slightly higher accuracy than the other two models.

Decision Trees are known for their ability to handle non-linear relationships and outliers, which may explain why it performed slightly better than the other two models.

RECOMMENDATION

Use the decision trees model to predict the functionality of water wells in tanzania.

Conduct regular monitoring and evaluation of the predictive model to ensure it remains accurate and effective in identifying functional wells.

FUTURE IMPROVEMENTS

- Feature engineering: consider adding additional features that could be relevant to the problem.
- Hyperparameter tuning: consider tuning the hyperparameters using techniques such as grid search or random search.
- Imbalanced data: consider using techniques such as oversampling the minority class or undersampling the majority class to balance the dataset (SMOTE).

CONTACT INFO

Email: zenah.biwott@student.moringaschool.com

Github: Biwott54

Linkedin: linkedin.com/in/zenah-biwott-416190221


Thank you !!!