CALVINE DASILVER

TANZANIA WATER WELLS

Predicting conditions of water wells with a Machine Learning Classifier

We urge the Tanzanian government to adopt our improved water point maintenance model, ultimately ensuring citizens' water needs are met.

Project Goal

BUSINESS UNDERSTANDING

- Nearly 24 million people in Tanzania struggle to find clean water.
- Water wells have emerged as a crucial solution to water scarcity, providing a reliable source for many communities.
- We need to develop a model to identify areas for improvement in well maintenance operations.

DATA UNDERSTANDING

- **Taarifa and Tanzanian Ministry of Water provided original data**
- **Data has 59,400 data points and 41 column**
- **will follow the OSEMN framework**

1. **Obtain**
2. **Scrub**
3. **Explore**
4. **Model**
5. **Interpret**

DATA CLEANING

1

We did Dimensionality Reduction. we dropped features that we won't use to model.

2

we checked for missing values and we did imputation to solve them.

3

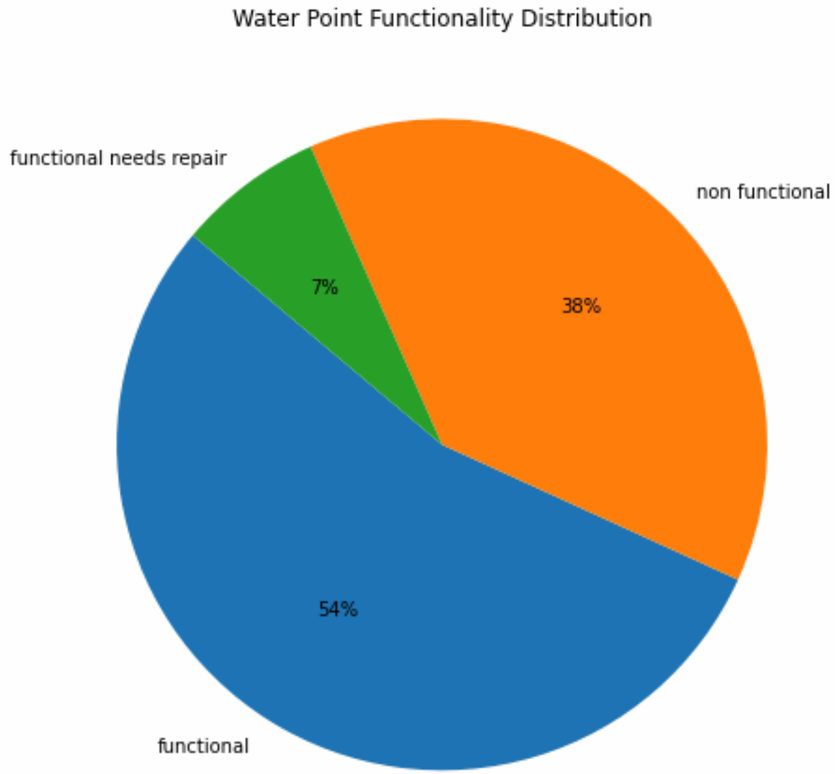
There were outliers in the data. We capped outliers to the upper bound of the interquartile range (IQR).

4

Lastly, we did a correlation using a heat mat to check if features were still correlating to each other, 75% and above.

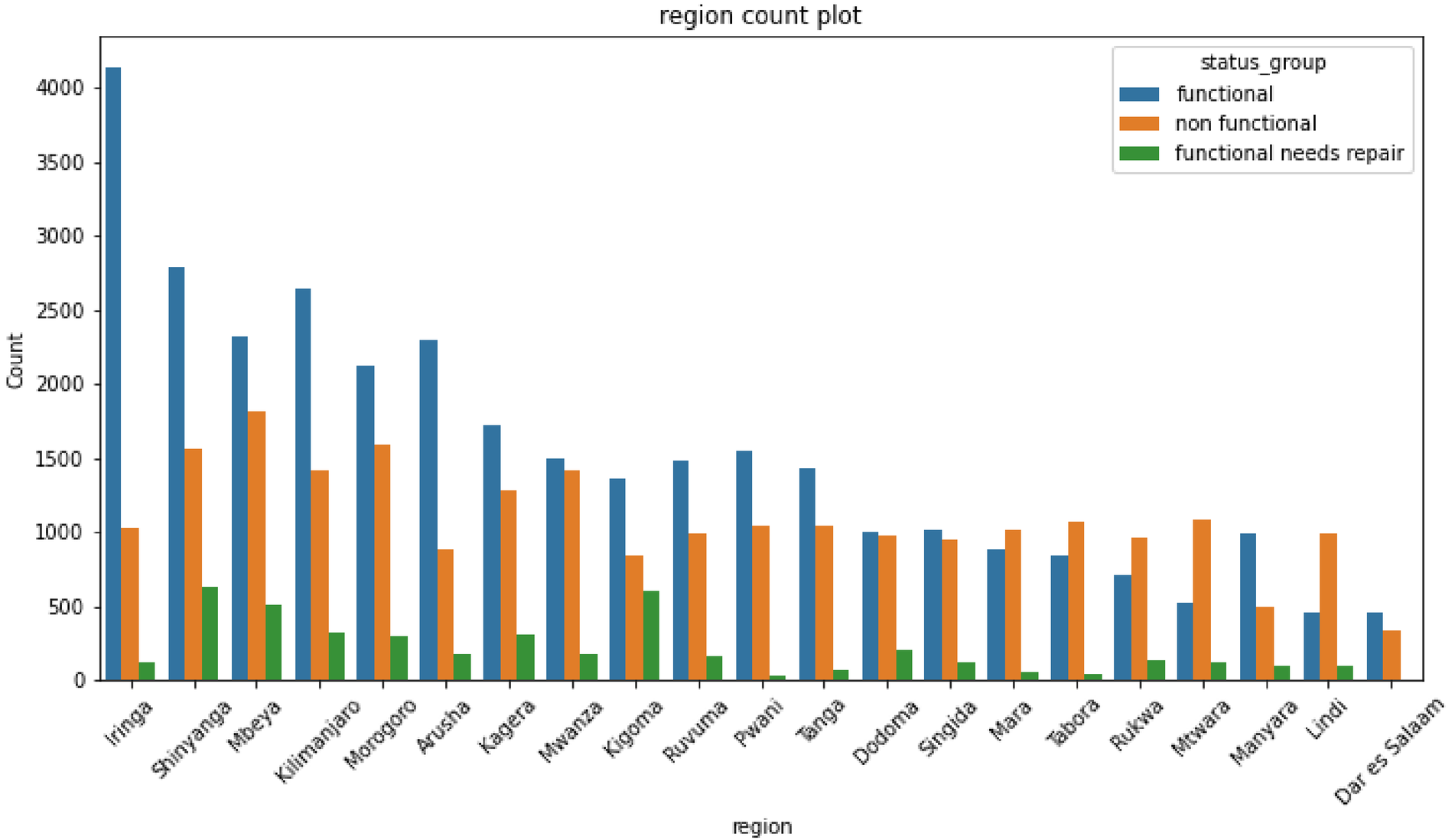
DATA VISUALIZATION

Target variable (status_group)



From our observation:

- 54% of water points are functional,
- 38% are non-functional
- 7% are functional and need repair



Analysis shows regional pump functionality varies. Iringa, Shinyanga, and Kilimanjaro have the most functional pumps, while Kilimanjaro and Morogoro have the most non-functional. Kigoma has the most functional pumps needing repair, suggesting targeted maintenance efforts there.

Clean water changes absolutely everything.

When a community gets access to clean water, it can change just about everything. It can improve:

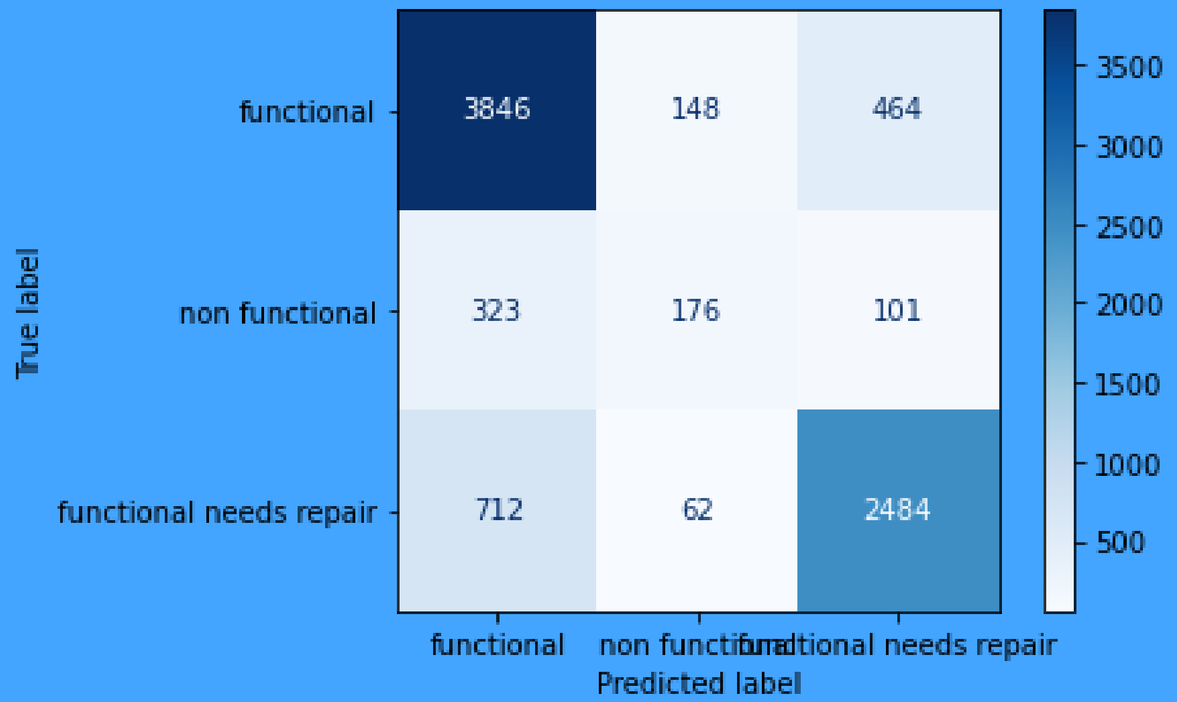
The

Access to food

Local economy growth

Education

RandomForestClassifier model

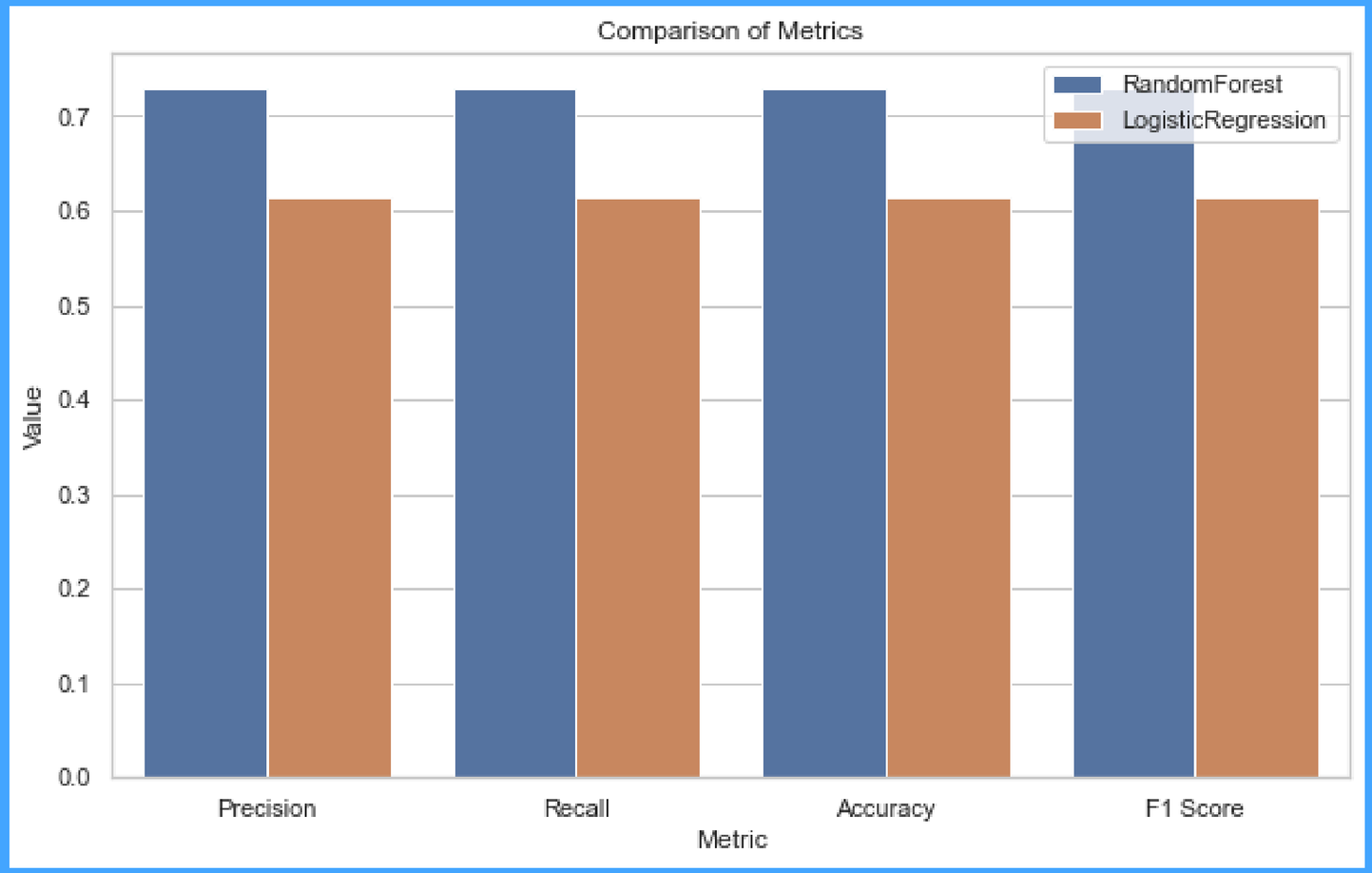


RandomForestClassifier model report
Validation accuracy: 78%

Streamline Maintainance and Repairs

Use Funding Efficiently and Effectively

Model Selection



- **Developed 4 different classifier based on the metrics.**
- **Out of the 4 RF was our best model while LR had the lowest Accuracy.**

The Classifiers are:
RandomForest
LogisticRegression
KNN
XGboost

Future Improvement

IMPROVE DATA

Enrich the model with coded qualitative data

MONITOR WELLS

Continuously improve the model to predict maintenance needs for well pumps

GEOGRAPHIC REGION

the model has to consider regional factors like rainfall,climate etc.

Clean water changes absolutely everything.

When a community gets access to clean water, it can change just about everything. It can improve:

THANK YOU