WATERPOINT OPERATIONAL STATUS PREDICTION PROJECT

ENSURING CLEAN WATER ACCESS IN TANZANIA

ISAAC KINYANJUI NGUGI

29TH SEPTEMBER 2024

Table Of Content

Overview

What is the Project About?

We aim to predict whether waterpoints in Tanzania are working well, need repairs, or are not functioning at all.

Why is it Important?

Access to clean water is vital for health and well-being. This project helps ensure that communities have reliable water sources.

Business and Data Understanding

Who Benefits?

Local communities, the Tanzanian Ministry of Water, and NGOs involved in water projects.

What Data Do We Use?

Information about waterpoints, such as their location, type, age, and condition, collected from various sources.

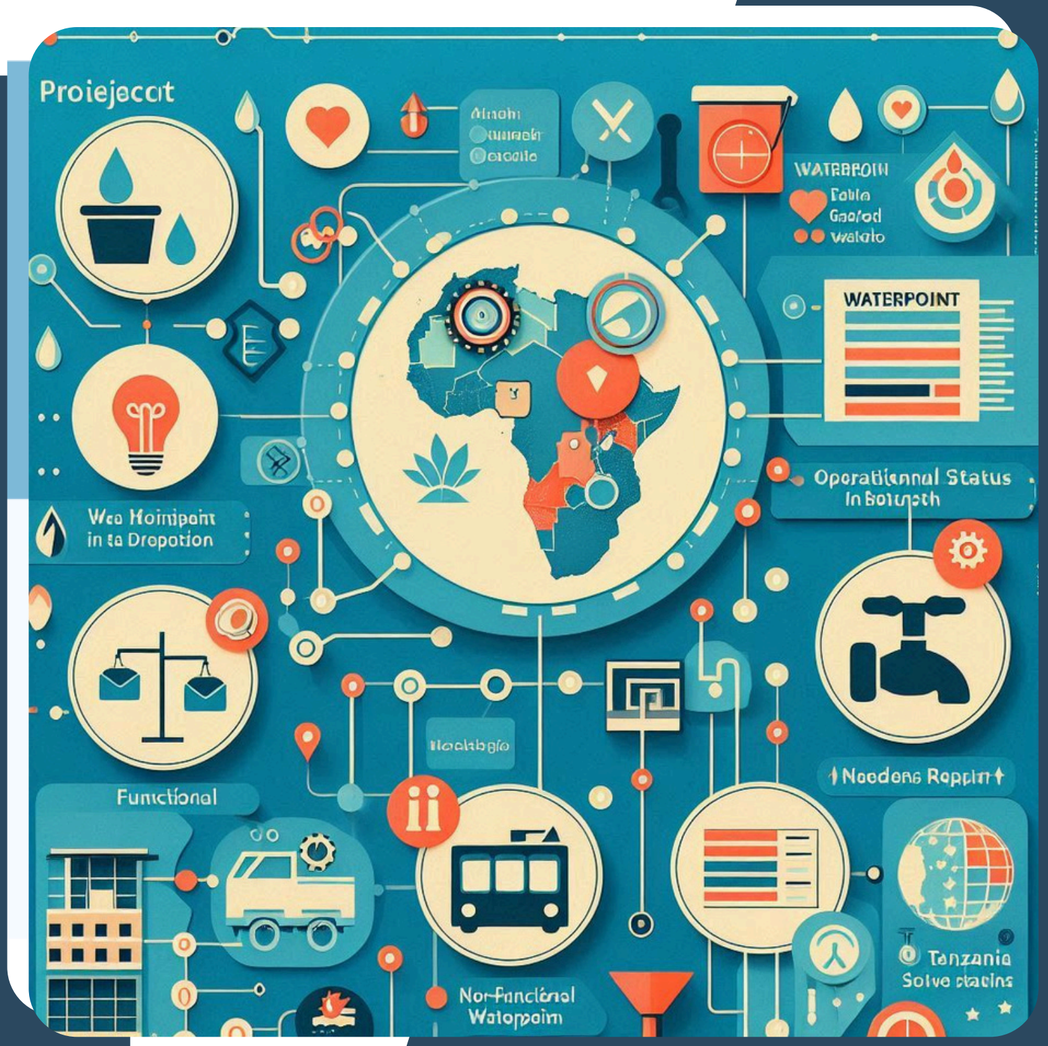
Evaluation

How Do We Know Our Model is Good?

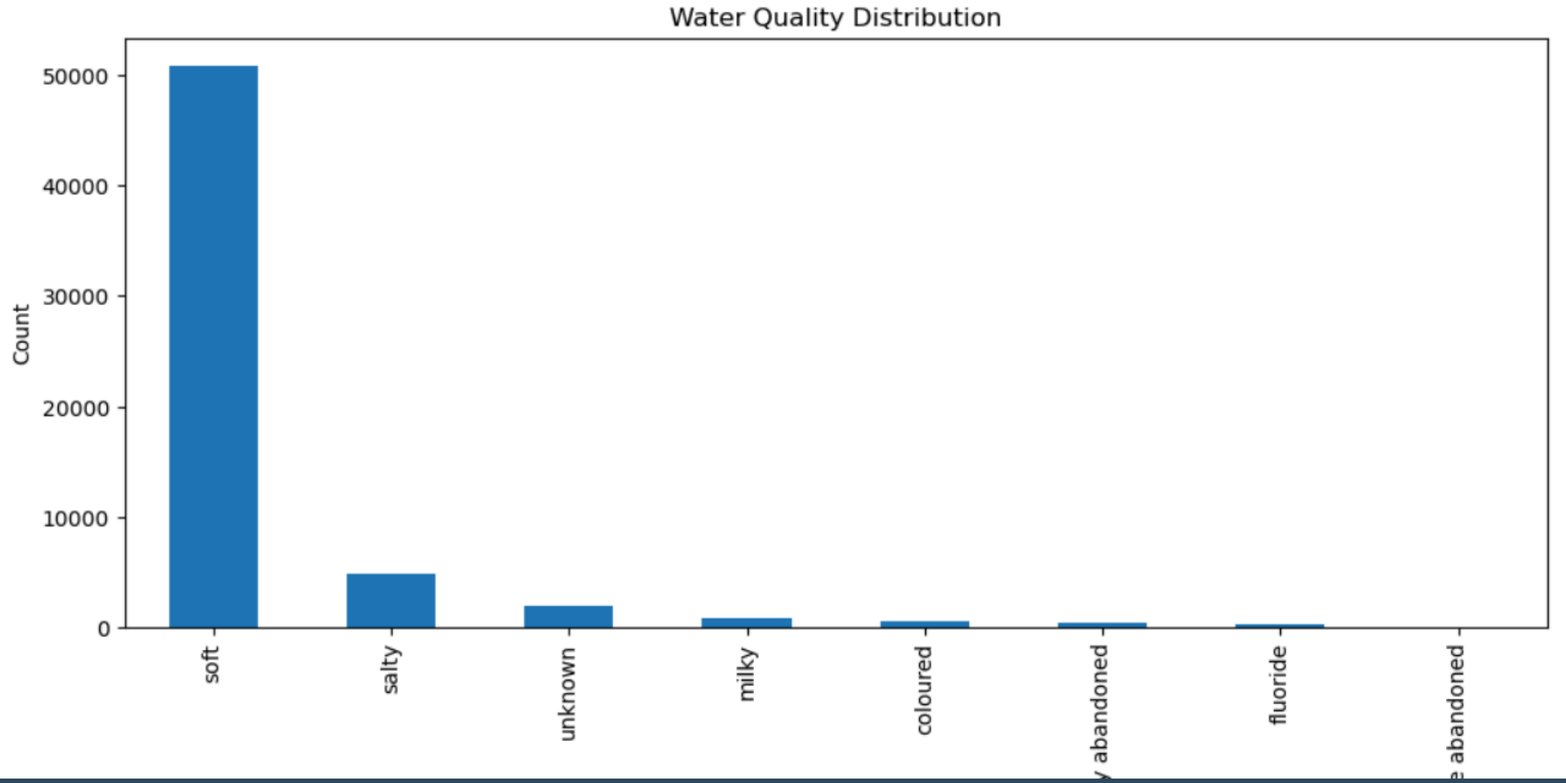
We check how often our predictions are correct. This is like a teacher grading a test.

Key Metrics:

- Accuracy: The percentage of correct predictions.
- Confusion Matrix: A simple table that shows how many times we got it right and wrong for each status.



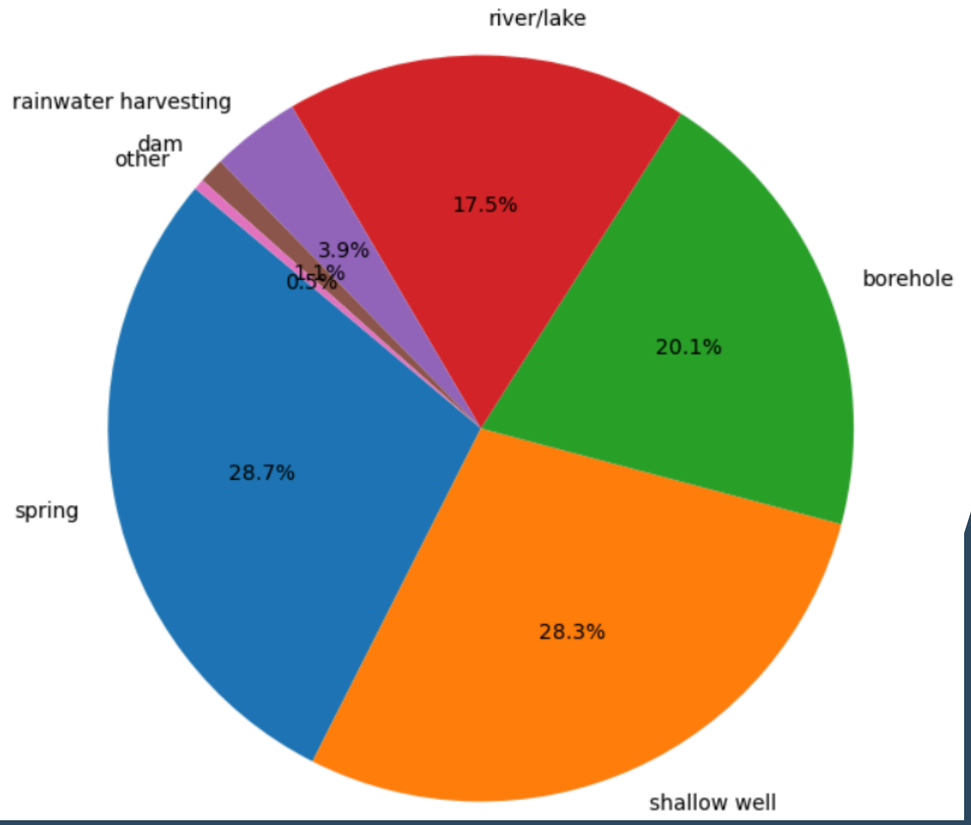
Water Quality Distribution



A bar chart displaying the distribution of water quality levels.

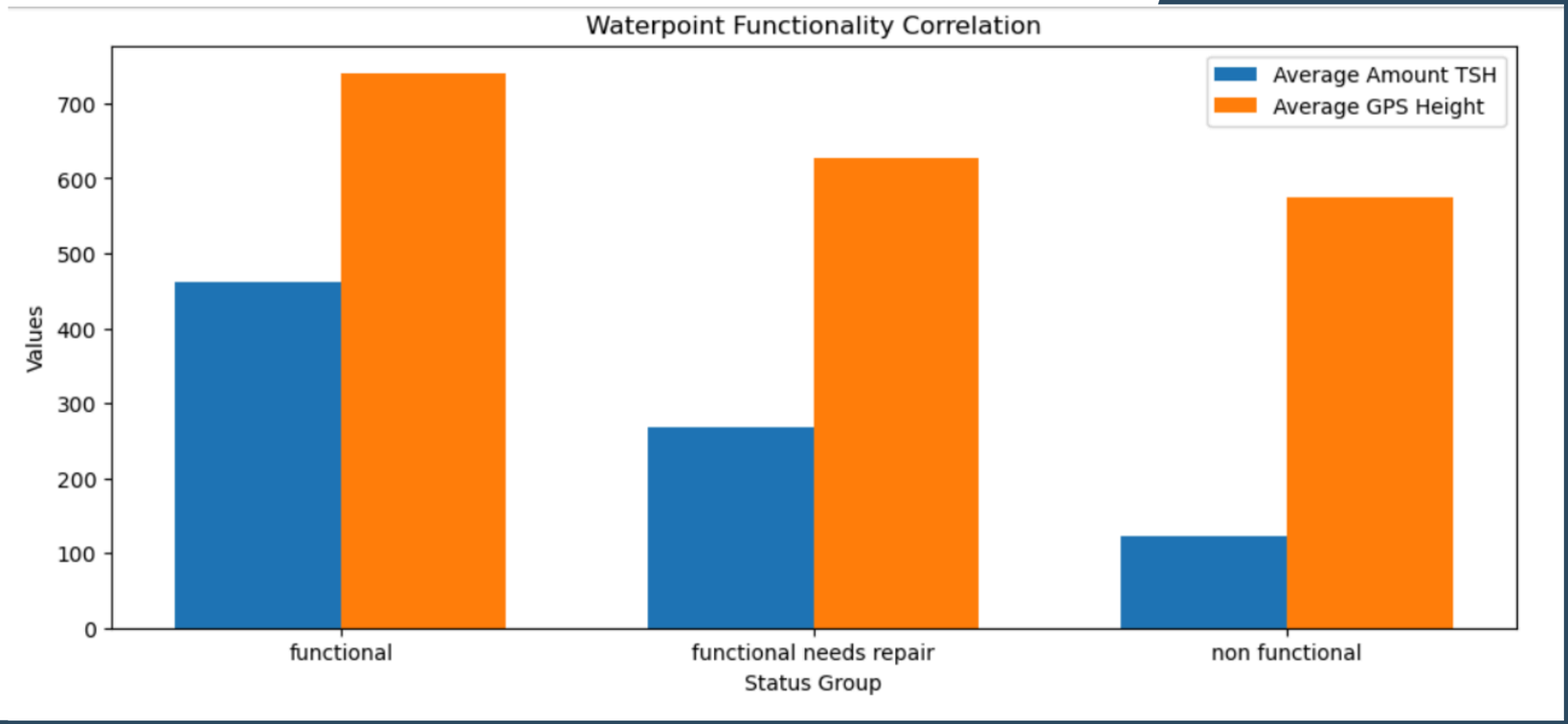
- Soft Water: This category has the highest count by a significant margin, indicating that most waterpoints have soft water quality.
- Other Categories: Salty, Unknown, Milky, Coloured, Salty Abandoned, Fluoride, and Fluoride Abandoned have much lower counts, with some categories having almost negligible counts.

Source Type Prevalence



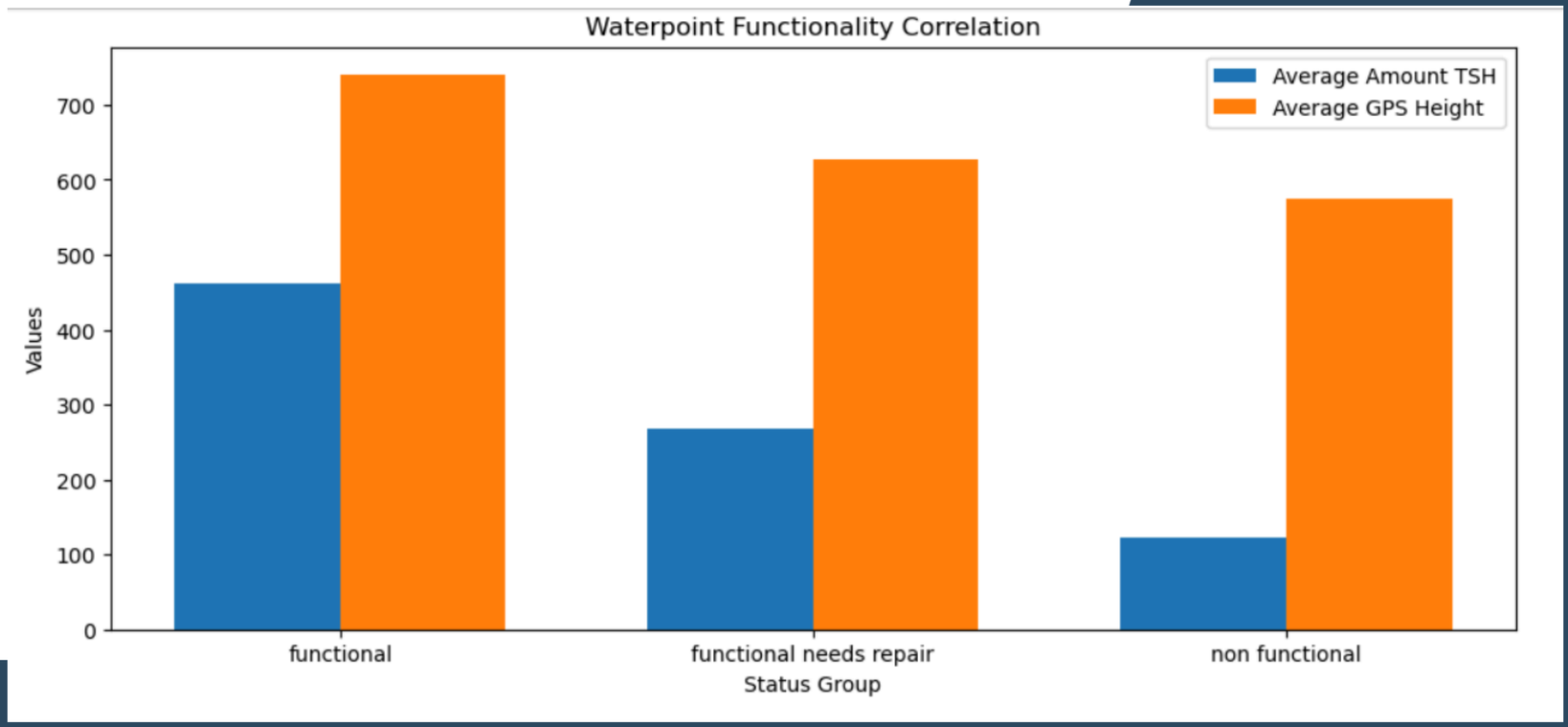
- **Dominant Sources**: Spring and Shallow Well are the most prevalent water sources, each accounting for nearly 29% of the total.
- **Moderate Sources**: Borehole and River/Lake are also significant, making up 20.1% and 17.5% respectively.
- **Least Common Sources**: Rainwater Harvesting and Dam/Other are the least common, with Rainwater Harvesting at 4.9% and Dam/Other at just 0.5%.

Waterpoint Functionality Correlation:



- **Functional Waterpoints**: Both average amount TSH and average GPS height are moderate, with TSH being slightly higher.
- **Functional Needs Repair**: Significant increase in both average amount TSH and GPS height, with TSH showing a notably larger increase.
- **Non-Functional Waterpoints**: Dramatic decrease in both values, especially for average amount TSH, which drops to nearly zero, while GPS height maintains a small value.

Age of Waterpoints vs. Status



- **Newer Waterpoints (Age 0)**: Predominantly functional, indicating that newer installations are mostly operational.
- **Older Waterpoints (Age 1750)**: A mix of statuses, with significant portions being functional, non-functional, and needing repair. This suggests that older waterpoints have a varied operational status.

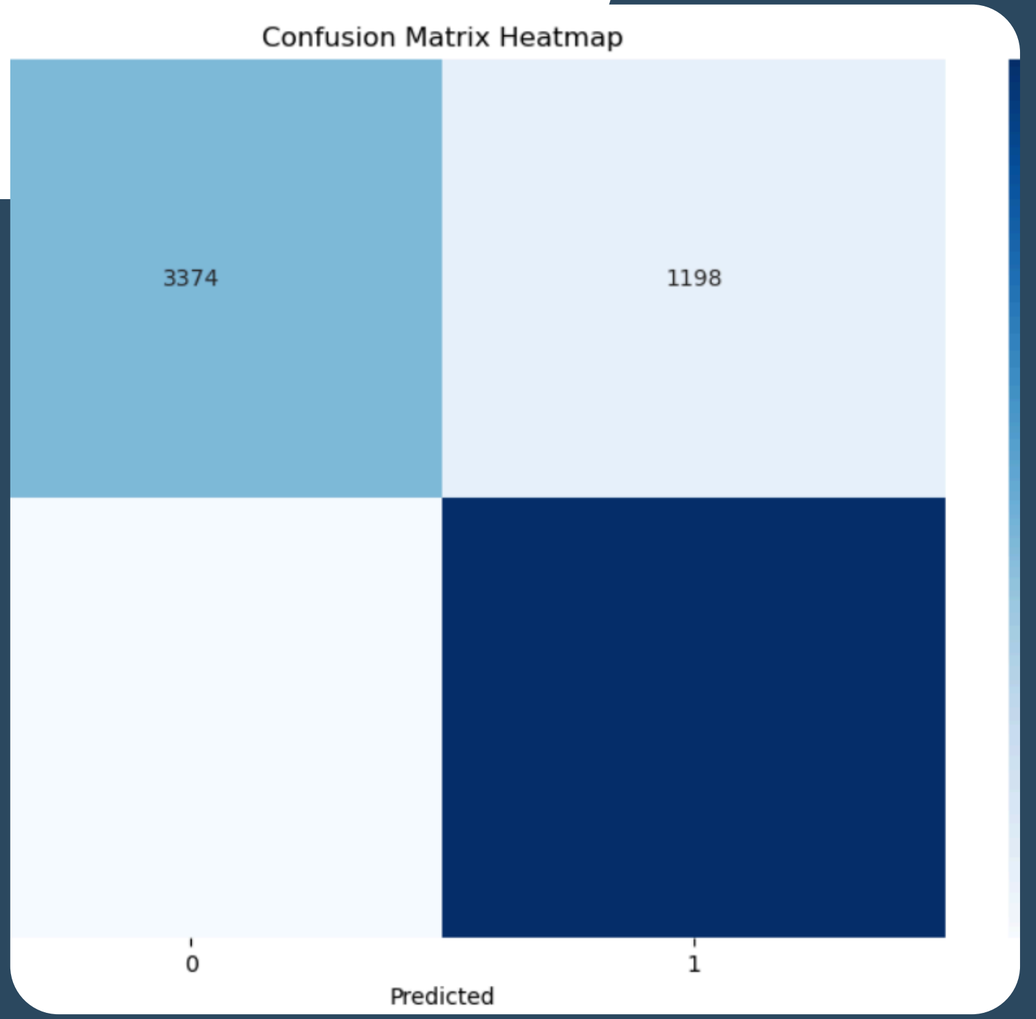
Model

K - NEAREST NEIGHBOUR

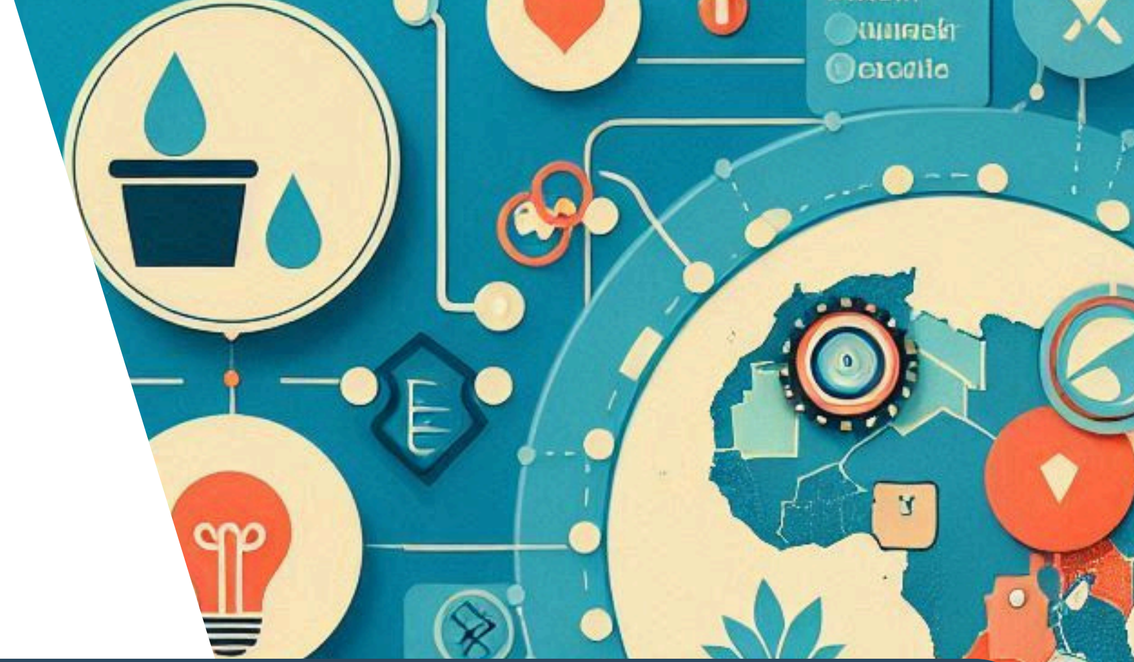
- **High True Positives and True Negatives**: The model correctly identifies a large number of both positive and negative instances, contributing to its overall high accuracy.

- **False Positives and False Negatives**: There are some misclassifications, with 1198 false positives and 819 false negatives. However, the number of true positives and true negatives is significantly higher, indicating strong model performance.

- **Precision and Recall**: The high precision (0.84) and recall (0.89) for the positive class suggest that the model is effective at identifying positive instances with few false positives and false negatives.



Recommendations



What Should We Do Next?

Focus on maintaining waterpoints that are predicted to need repairs.

Use our findings to guide future waterpoint installations in areas with high demand.

Engage Communities: Involve local communities in monitoring and reporting the status of waterpoints.

Next Steps

01.
Solution

Implement Predictive Maintenance: Use our model to prioritize repairs.

02.
Solution

Expand Data Collection: Gather more data to improve predictions.

03.
Solution

Train Local Teams: Educate local maintenance teams on using the model for better decision-making.

THANK YOU

Isaac Kinyanjui Ngugi
Junior Data Scientist

Questions?

Thank the audience for their time and invite any questions or discussions.