Beyond the Surface:

Leveraging Machine Learning to Monitor Rural and Urban Wells Functionality in Tanzania to Promote Access to Basic Water.

Vanessa Sandra

Agenda

- Introduction
- Project Overview
- Data
- Approach
- Key Insights
- Recommendations
- Q&A Session
- Conclusion

Introduction

For the longest time, access to basic water has been a challenge in Tanzania. Despite the government's efforts to solve this problem, Tanzania still remains a water-stressed nation.

In 2006, the Tanzanian government adopted a National Water Sector Development Strategy that aimed to promote integrated water resources management and the development of urban and rural water supply.

However, almost 2 decades later, **only 60% of Tanzanian households have access to basic water**. This project investigates why and offers **solutions to help the Water Sector Development Strategy achieve 100% water access** in the country over the next decade or less.

Project Overview

Problem Statement

"1 in 4 wells in Tanzania, is non functional or needs repairs"

Access to safe, working water wells is a basic need, yet many wells across Tanzania are broken, unreliable, or abandoned. Without regular monitoring or clear data, it's hard for the government or organizations to know which wells need repairs or why they're failing.

- This project uses machine learning to predict the status of water wells before they break down, helping decision-makers target repairs, use funds wisely, and keep water flowing where it's needed most.

Project Objective

The project seeks to:

- Develop a machine learning model that predicts a well's functionality status (functional, non functional, needs repair).
- Create a system that will help the Tanzanian government monitor wells across Tanzania.
- Identify features that drive high durability in wells.
- Come up with recommendations that will help the Water Sector Development Strategy maximize their operations to serve the nation better.

Data

Data Source

We used real-world data sourced from Taarifa, an open-source platform that aggregates data from the Tanzania Ministry of Water and the Tanzanian Ministry of Water.

Data Cleaning

In order to have reliable data, we followed the best practice to clean our data, this involved:

- Handled missing values.
- Standardized texts.
- Dropped irrelevant features.

Data Preprocessing

To prepare for effective modeling, we preprocessed the data by:

- Encoding – turning texts to numerical.
- Balanced classes to prevent biasness.

Approach

To understand why some water wells stop working — and how to prevent it — we analyzed thousands of well records from across Tanzania using data science. Here's how we approached the problem:

- Explored the Data for Patterns
  We looked closely at the data to find patterns. For example, we discovered that wells built by certain funders or installed using specific methods were more likely to last longer.
- Trained a Smart Model
  We used machine learning, a method that teaches the computer to find patterns and make predictions, to build a model that can guess whether a well is:
  - Fully functional
  - Not working
  - Working, but in need of repair
- Improved the Model Over Time
  We tested different versions of the model and made improvements until we found one that gave us the most accurate predictions.
- Learned What Matters Most
  The model also told us which factors were most important. For example, it highlighted that wells built using gravity-based systems and natural springs tend to last longer.

Modeling Approach

This was the first model we used; however, it **assumes the predictors have a linear relationship with the target variables**, our dataset had a non-linear relationship. It had a **prediction performance of 62%.**

Due to the previous models' shortcomings, we ended up using this ensemble classifier as it **generalizes data well and it mitigates overfitting, improves accuracy, and handle complex datasets more effectively**. It had a **prediction performance of 88%,** slightly better than the Decision Tree but well positioned to work with such a complex data.

Positioned to handle multiclass classification and non-linear relationships, this was better suited for this project. But it was vulnerable to overfitting. It had a **prediction performance of 86%.**

Evaluation Metrics

For evaluation, we used the following metrics:

- Accuracy – This measured how accurate our model was at predicting well functionality status.
- F1-score – This helped us know how our model was performing generally.

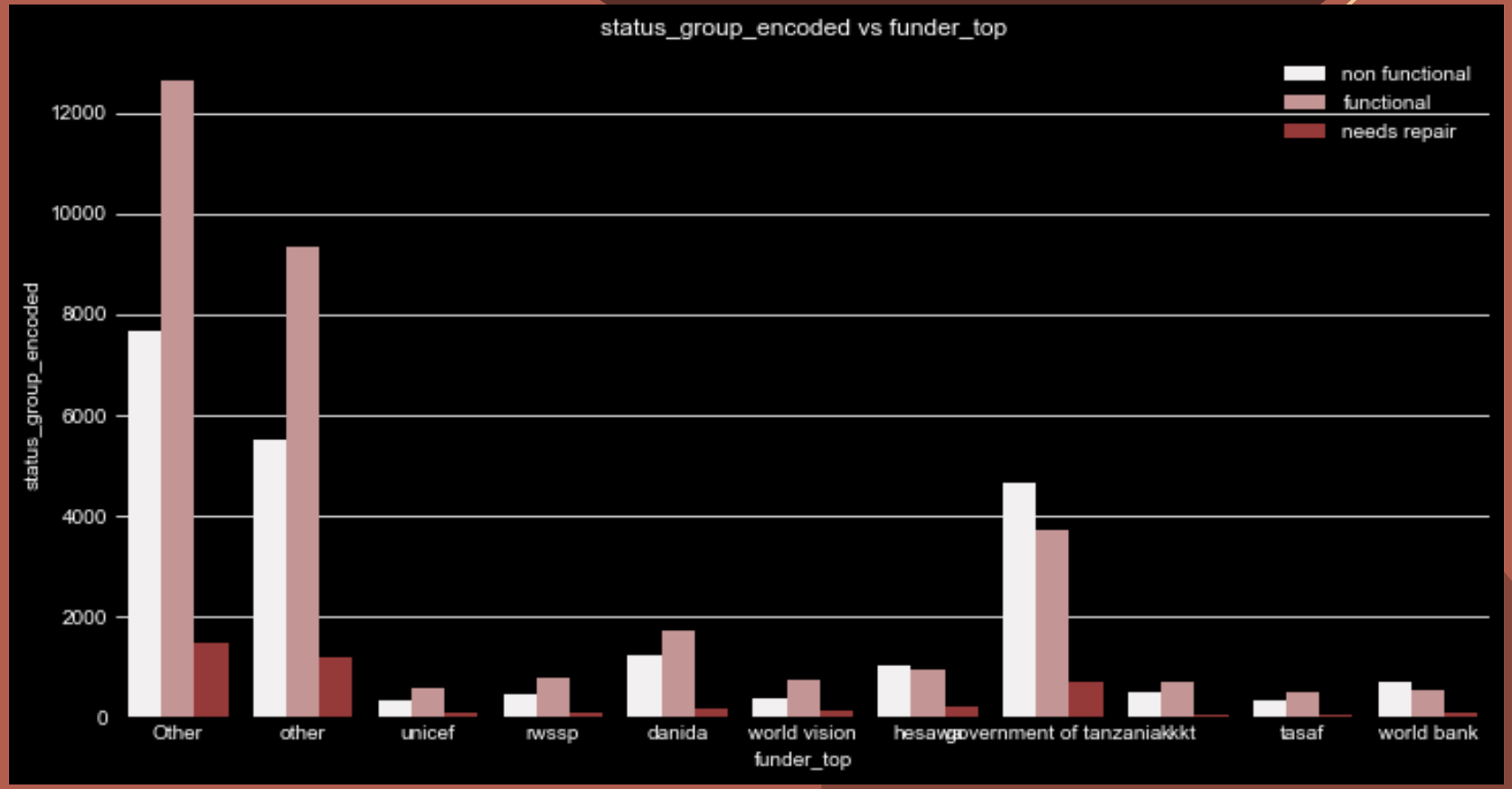
The table below shows us how the model performed on new data:

Accuracy	F1-Score
77	68

Key Insights

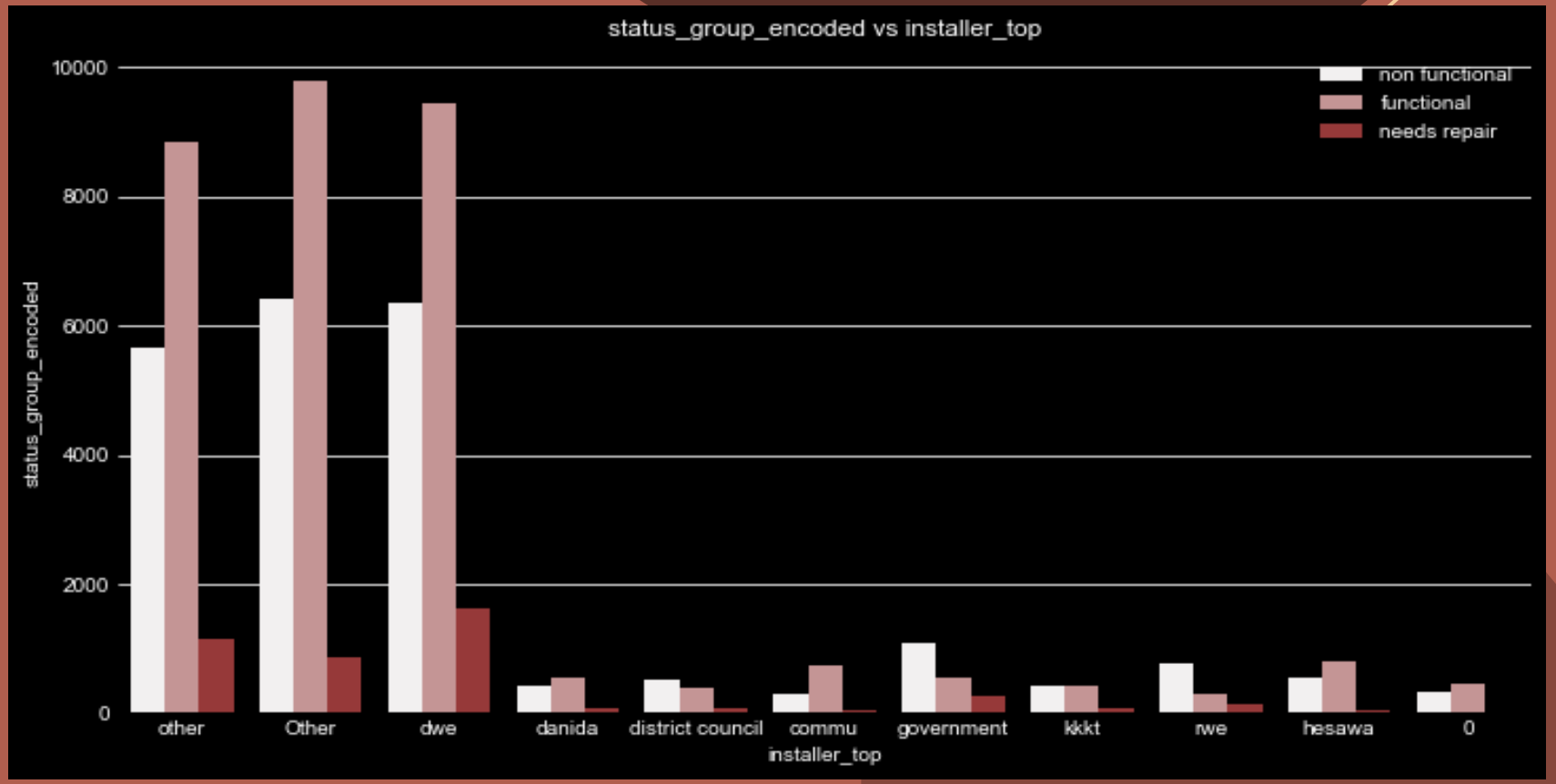
Funders

- Government-funded wells were the most common by single entities, but also had the highest failure rate – indicating that quantity alone does not ensure quality.
- Small organizations, these include churches, schools, mosques, have contributed the most towards promoting water accessibility across Tanzania.



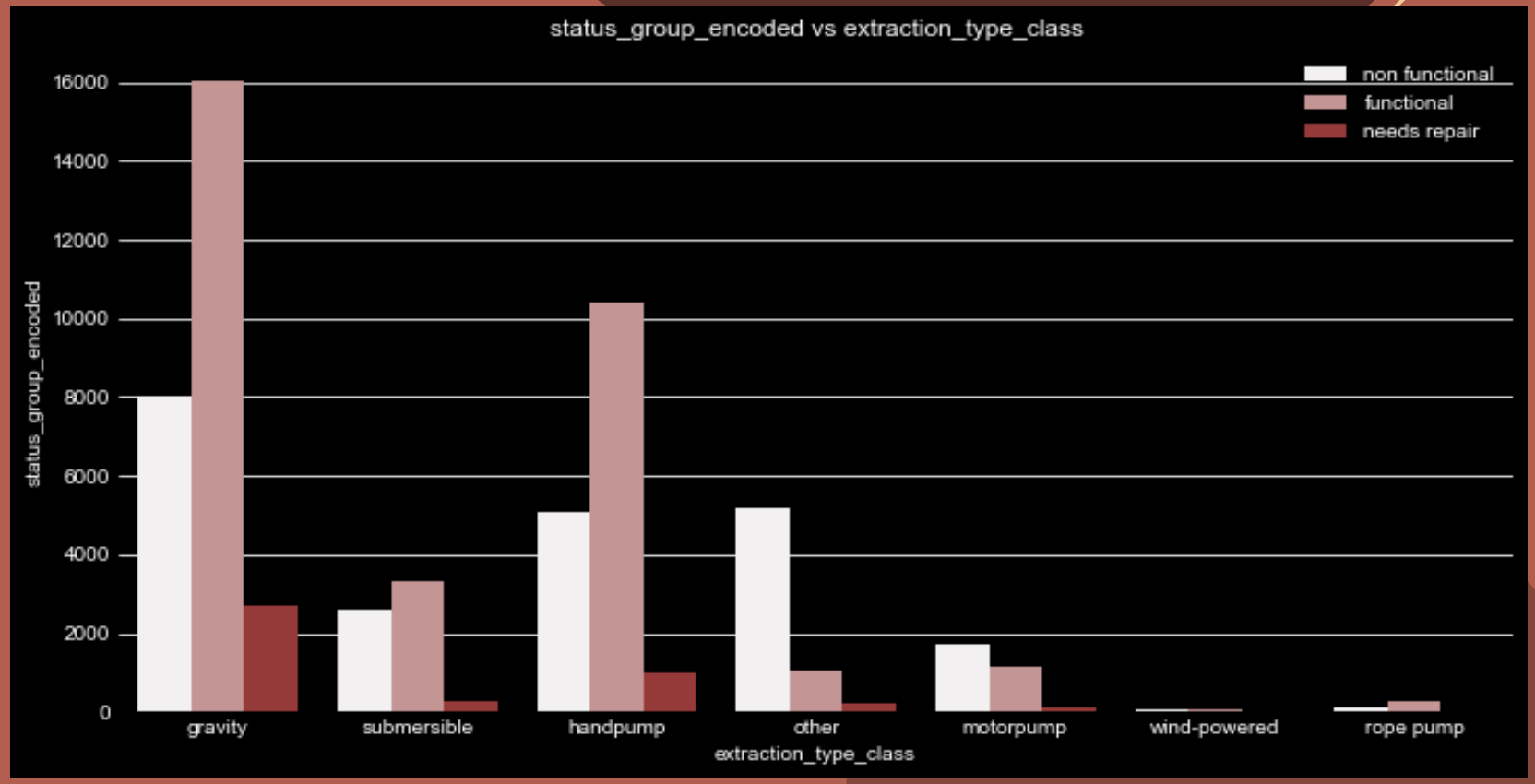
Installers

- Wells installed by the District Water Engineer had the longest-lasting functionality, highlighting the importance of technical expertise.
- Wells installed by the communities also had longer-lasting functionality, in contrast to the ones built by the government.



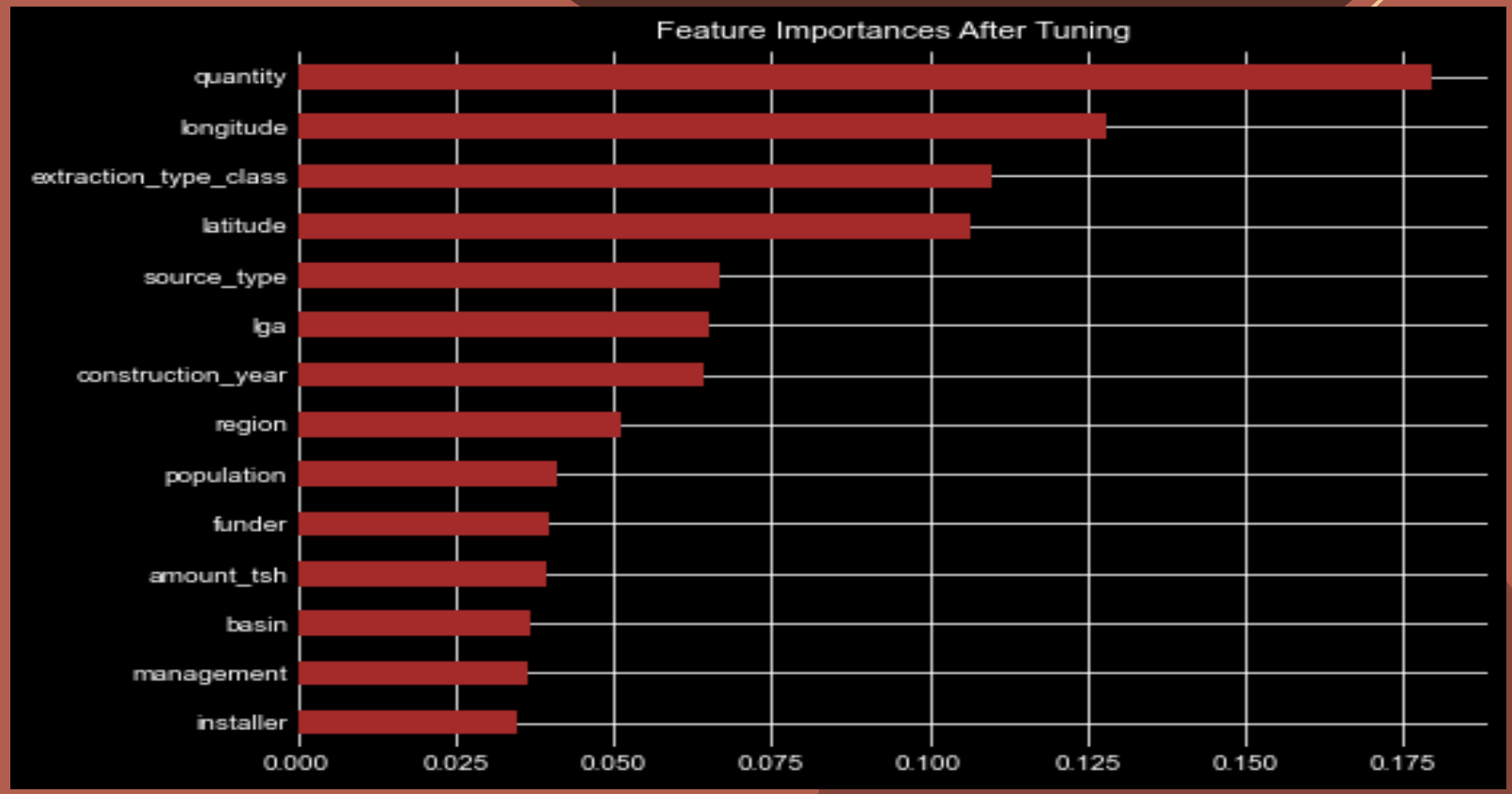
Extraction Types and Water Sources

- Gravity-fed and hand pump extraction methods were the most effective extraction types, and had the highest numbers of functioning wells.
- Additionally, spring-based water sources were strongly associated with longer-lasting well performance.



Important Features

- Certain features such as extraction type, installer, region, and source type played a significant role in predicting well functionality.



Recommendations

- Enhance government accountability in project execution.

We recommend that the Tanzanian government establish monitoring frameworks for government-funded wells to ensure installations follow quality assurance protocols and receive proper maintenance.

- Standardize installations through certified professionals.

We recommend that all wells, especially those funded by public or NGO sources, be installed under the supervision of certified District Water Engineers.

- Adopt gravity-based extraction systems wherever viable.

Given their durability and low maintenance, gravity-fed wells should be prioritized, particularly in regions where topography supports them.

- Prioritize spring sources during site selection.

Water source plays a critical role in the longevity of a well. Therefore, we recommend that water sourcing should be more deliberate. Feasibility studies should assess whether natural springs are available and accessible before selecting a site.

Summary

Machine Learning can enhance water access, and while further improvements are always possible, the current model provides a solid foundation for decision making and resource allocation, and represents a strong first step towards sustainable water access using data science.

The next step is deployment, with the current model, the Tanzanian government can begin monitoring their past projects, allocating resources to places needed and start planning their next projects.

"Water is just not a basic need, water is LIFE!"

Q&A Session

Thank You

Vanessa Sandra

veesandra30@gmail.com

+254 706 894 535