PREDICTING WATER WELL FUNCTIONALITY IN TANZANIA

Godfrey Osundwa

PROJECT OVERVIEW

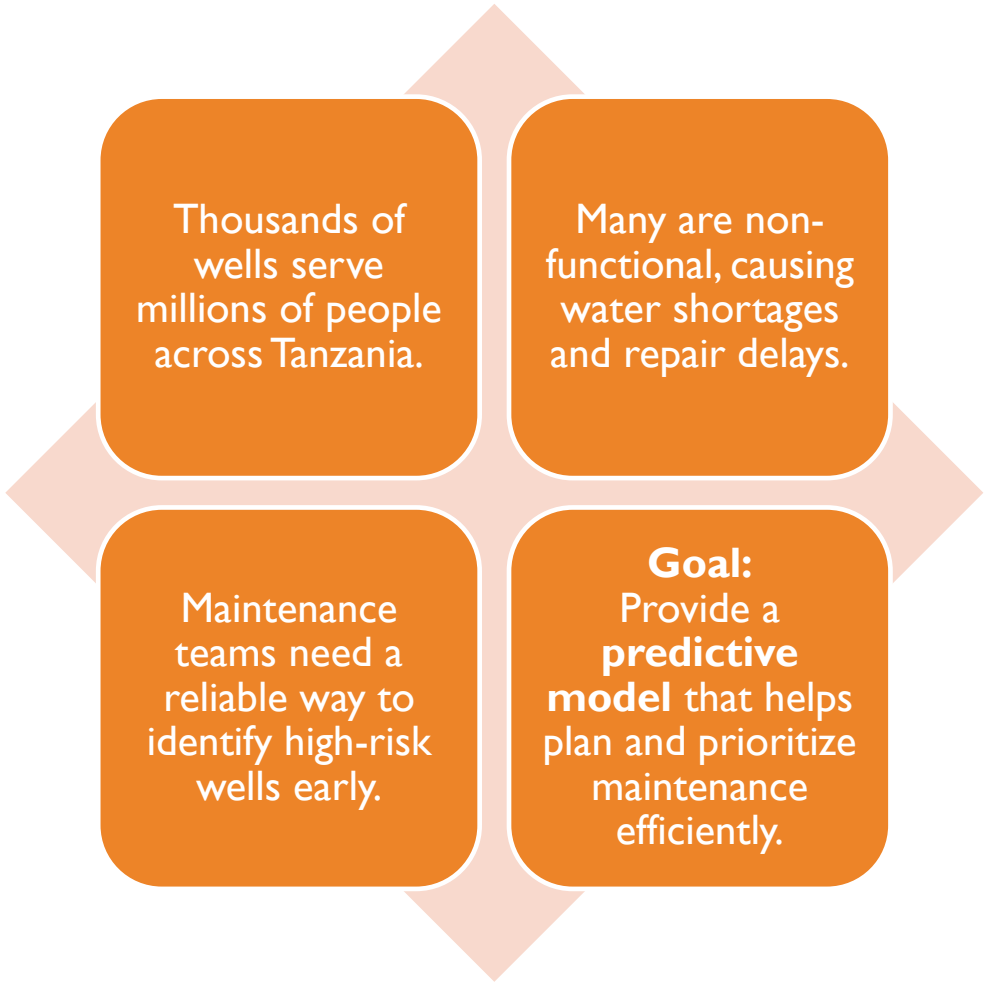
Why This Project Matters

Access to clean water is essential for life and development.
In Tanzania, many wells stop working due to poor maintenance or aging infrastructure.

This project uses data science to predict which water wells are **functional**, **non-functional**, or **need repair**, helping decision-makers prioritize resources and improve service delivery.

BUSINESS UNDERSTANDING

The Business Challenge



DATA UNDERSTANDING

About the Data

We used the **Tanzania Water Wells dataset** from the DrivenData "Pump It Up" competition.

It includes data on **59,000+ wells**, such as:

Location: region, district, coordinates

Construction: installer, year built, water quantity

Management: who runs the well, how users pay

Target: current condition — functional, needs repair, or non-functional

This data helps uncover what factors contribute most to well performance.

MODELING APPROACH

How We Built the Model

We used **classification modeling**, which means the model predicts categories (e.g., "Functional" vs "Broken").

Two models were developed:

Logistic Regression – A simple model to set the baseline.

Random Forest Classifier – A more advanced model capturing complex relationships.

Both models were trained, tested, and compared to find the best performer.

RESULTS

How Well Did the Models Perform?

Random Forest gave the highest accuracy and reliability.

It correctly predicted most wells' conditions and captured patterns missed by the simpler model.

Key takeaway:
The model can **accurately identify high-risk wells**, allowing proactive maintenance decisions.

KEY INSIGHTS

What Factors Matter Most

The data revealed several important patterns:

Construction Year: newer wells tend to work better.

Water Quantity: low or dry wells often fail.

Management Type: wells managed by professional organizations are more reliable.

Region: some areas experience higher failure rates due to environmental conditions.

These insights can guide where to invest in repairs and new infrastructure.

RECOMMENDATIONS

What We Recommend

PRIORITIZE MAINTENANCE ON OLDER WELLS AND THOSE WITH LOW WATER OUTPUT.

USE PREDICTIVE MODELING TO CREATE A NATIONAL MAINTENANCE PLAN.

FOCUS FUNDING ON HIGH-RISK REGIONS.

TRAIN LOCAL TEAMS TO COLLECT BETTER MAINTENANCE AND MANAGEMENT DATA.

NEXT STEPS

What Comes Next

INTEGRATE THE MODEL INTO A **MOBILE OR WEB TOOL** FOR FIELD TEAMS.

COLLECT MORE **REAL-TIME WELL DATA** TO IMPROVE ACCURACY.

PARTNER WITH GOVERNMENT AND NGOS FOR **NATIONWIDE DEPLOYMENT**.

CONTINUE REFINING THE MODEL FOR LONG-TERM SUSTAINABILITY.

THANK YOU

Presenter name: Godfrey Osundwa

Email address: godfrey.osundwa@student.moringaschool.com