Tanzania Water Wells

Henry Njoroge

CONTENTS

01

Motivation & Goal

Why Wells Fail Too Often

Current State of Wells

In Tanzania, thousands of boreholes are idle or deliver unsafe water, wasting donor funds and eroding health gains. This infrastructure failure deepens rural poverty and forces communities to rely on long walks for water.

Impact on Communities

Broken water wells disproportionately affect rural areas, where access to clean water is critical for health, education, and economic productivity. Proactive failure prediction is essential to ensure sustainable water services.

Project Motivation

The project aims to shift from costly reactive inspections to predictive maintenance, leveraging data-driven insights to keep water flowing and directly support the United Nations Sustainable Development Goal 6 on clean water and sanitation.

Project Objective Defined

Objective Overview

The primary objective is to build an interpretable classifier that labels every well as functional, functional-needs-repair, or non-functional, enabling efficient maintenance planning and resource allocation.

Expected Outcomes

The project aims to surface key drivers of well failure through exploratory data analysis, providing actionable insights for policymakers and stakeholders to make informed decisions and reduce downtime.

02

Data & Features

Tanzania Water Point Corpus

Dataset Overview

The dataset from the Tanzania Ministry of Water includes over 59,400 water points with 39 attributes, covering geospatial, technical, and administrative details. This rich dataset forms the basis for predictive modeling.

Geospatial Data

Attributes such as GPS coordinates, altitude, and geographic water basins provide spatial context, helping identify regional patterns in well functionality and failure.

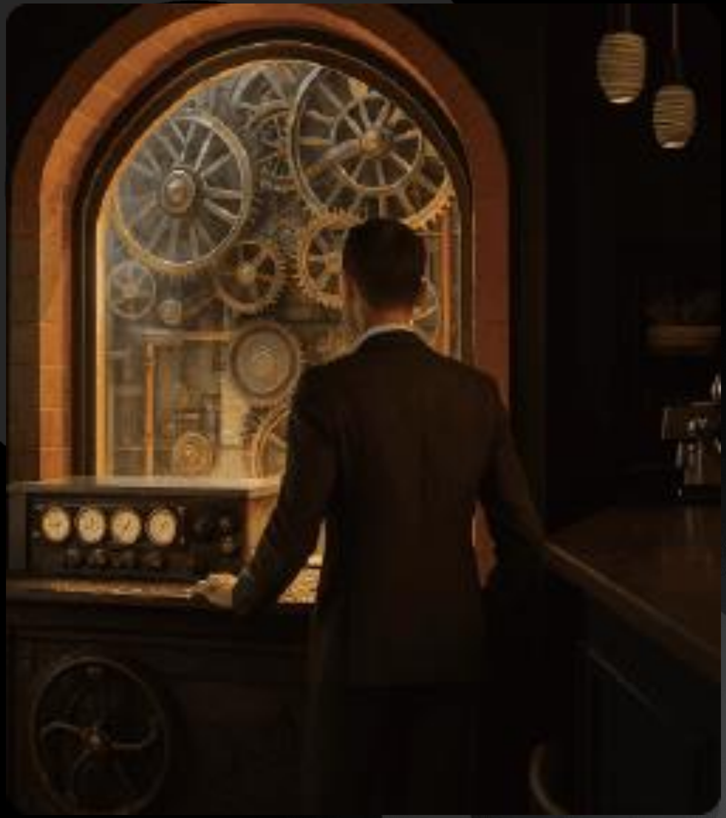
Technical Specifications

Technical attributes like extraction type, construction year, and water quantity offer insights into the hardware and operational aspects of each well, crucial for predicting maintenance needs.

Administrative Details

Administrative data, including funder, installer, and management model, reveals the organizational and regulatory context of water points, influencing their long-term sustainability.

Feature Groups for Modeling



Feature Groups

The core feature groups include location metadata, hardware specifications, administrative context, and water service indicators. These groups capture the multifaceted factors influencing well functionality and failure.

03

Model

Predictive models

Model Selection

We trained four supervised classifiers: Logistic Regression for baseline transparency, Decision Tree for rule clarity, Random Forest for bagged stability, and XGBoost for gradient-boosted power, all aiming for an F1 score above 0.65.

Model Comparison

Each model was evaluated on identical pre-processed data folds, with metrics like precision, recall, and confusion matrix analysis used to assess performance across the three functionality classes.

Training & Validation Protocol

Handling Class Imbalance

Class imbalance was addressed using SMOT (Synthetic Minority Over-sampling Techniqu which helps balance the dataset and ensures that the model can accurately predict minori classes like non-functional wells.

Hyperparameter Tuning

Hyperparameters were optimized using Bayesian search, allowing for efficient exploration of the parameter space and improving model performance through iterative tuning.

04

Deployment & Impact

Dashboard to Field Action

Interactive Dashboard

An interactive map filters wards by predicted failure probability, allowing maintenance teams to prioritize high-risk wells. This tool converts model outputs into actionable field assignments.

Impact on Maintenance

The dashboard tracks post-repair status updates, providing real-time feedback on intervention effectiveness and enabling continuous improvement in maintenance planning.

Next Steps & Scale Path

Integration into Workflows

The model will be embedded into the ministry's Computerized Maintenance Management System (CMMS) to streamline maintenance scheduling and resource allocation.

Capacity Building

Securing local data science capacity ensures long-term sustainability and enables continuous improvement of the model through local expertise.

Real-Time Data Integration

Future phases will incorporate real-time sensor data to enhance predictive accuracy and enable proactive maintenance before failures occur.

Regional Expansion

The approach will be replicated in neighboring countries like Rwanda and Uganda, extending the benefits of predictive maintenance and supporting regional water infrastructure sustainability.

THANK YOU