PREDICTING WATER WELLS

Data-Driven Insights for Smarter Water Management

PROJECT OVERVIEW

Clean water access is **vital for** community health and development, yet many wells in Tanzania fail due to **poor maintenance** or aging, making our machine learning predictions crucial.

PROJECT GOALS

- Enhance water quality monitoring
- Predict well functionality trends
- Support resource allocation efficiently

TARGET AUDIENCE

- Government agencies for policy-making
- NGOs focused on community health
- Local communities benefiting from clean water

DATA UNDERSTANDING

The dataset provided consists of information collected from thousands of water wells across Tanzania. The goal is to understand the structure, content, and quality of the data before any preprocessing or modeling is done.

FILES PROVIDED

- **Training Set Values : Contains the features for each water well, such as location, pump type, construction year, etc.**
- **Training Set Labels Contains the target label for each well—its operational status.**
- **Test Set Values: Unlabeled dataset with the same structure as the training set values, used to test our trained model.**

GOALS

- **Load and inspect the data.**
- **Understand the size, types, and distributions of columns.**
- **Check for missing values and inconsistencies.**
- **Merge the training values and labels for exploratory analysis.**

UNDERSTANDING CLASSIFICATION

- A method for sorting data into categories.
- The model identifies patterns for predicting outcomes.
- Enables quick and accurate decisions using data.
- Essential for analyzing large datasets effectively.

OUR MODELING APPROACH

We tested multiple models for accuracy.

Data was cleaned and balanced for fairness.

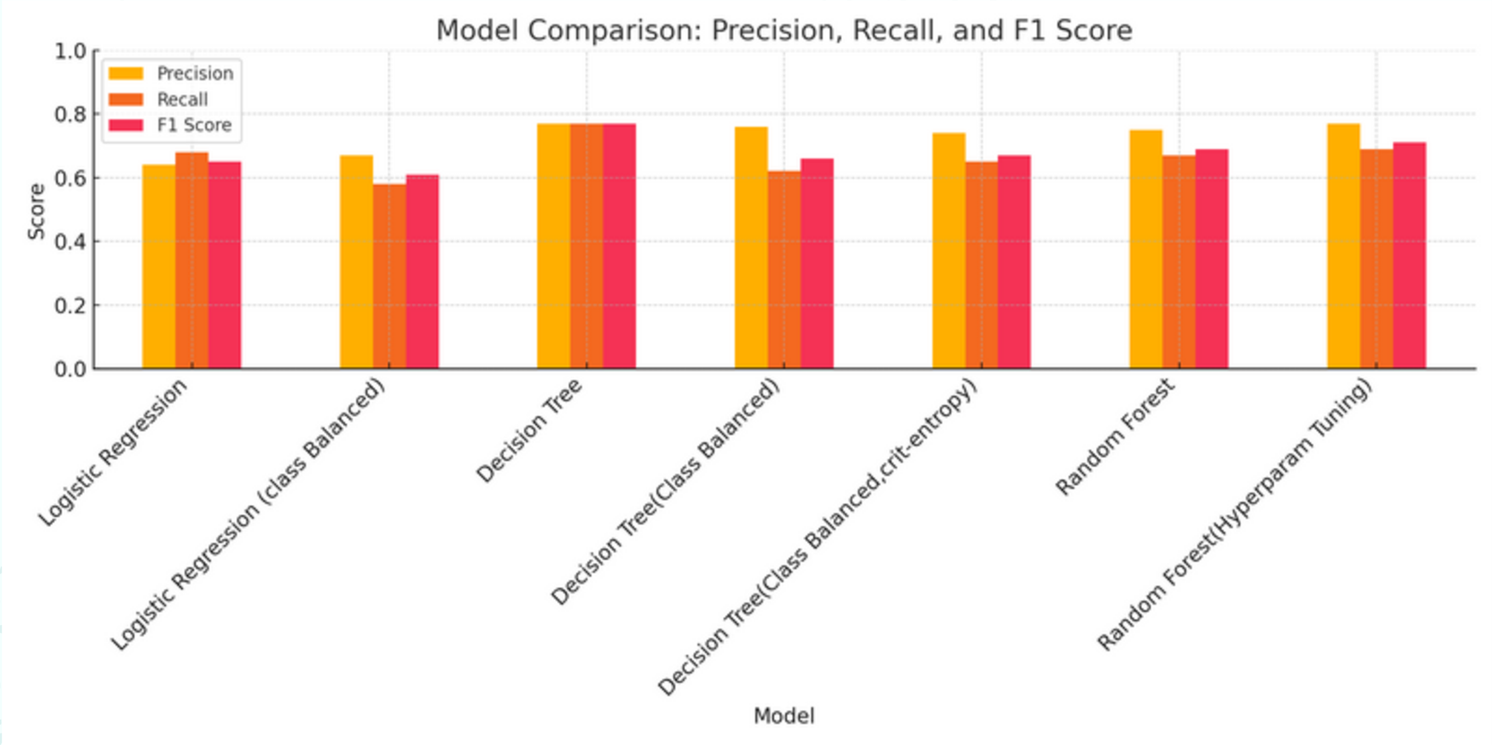
This improved predictions across all well categories.

Random Forest (a modelling technique) was chosen for high performance.

MODEL PERFORMANCE RESULTS

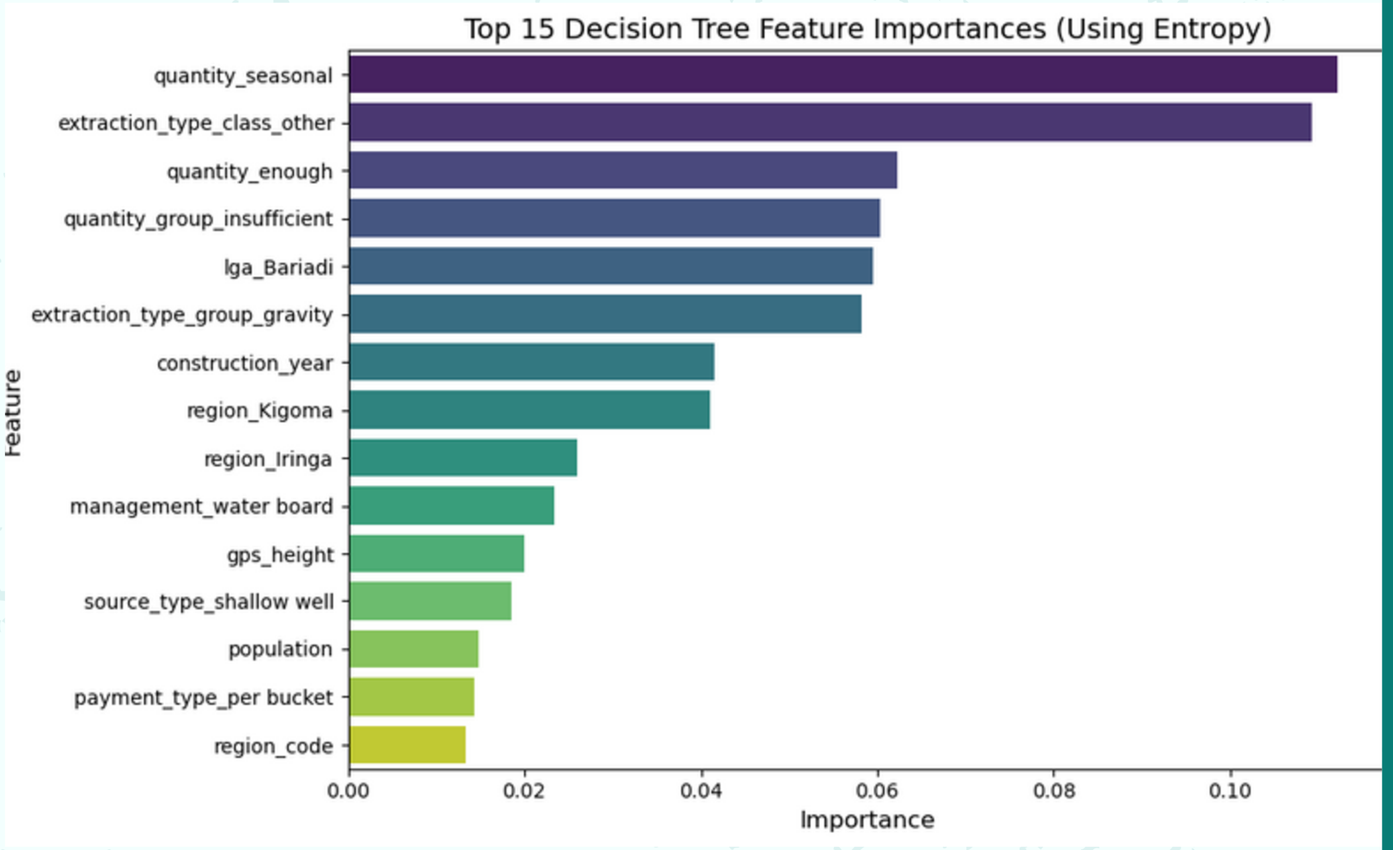
- Accuracy reached [69]% for predictions.
- F1 Scores: Functional: [0.71].
- F1 Scores:
  Needs Repair: [0.35],
  Non-Functional: [0.74].
- The model helps **target interventions** effectively.

F1 captures both precision (avoiding false alarms) and recall (not missing true cases). Accuracy alone would be misleading in this context.



KEY FACTORS INFLUENCING WELLS

- Construction year significantly affects well condition.
- Installer reputation impacts failure rates.
- Water quality is crucial for sustainability.
- Geographic location influences maintenance needs.



ACTIONABLE BUSINESS RECOMMENDATIONS

- Prioritize repairs in **high-risk areas.**
- Plan future installations based on **model predictions.**
- Avoid contractors with **poor performance history.**
- Save costs by focusing on the **right wells.**

SCALING THE SOLUTION

Integrating our model into decision-making will enhance water management and ensure efficient resource allocation across Tanzania's water wells.

IMPLEMENTATION STEPS

- Integrate model into current systems
- Use Tableau for live updates
- Train stakeholders on new tools

FUTURE EXPANSION

- Analyze other regions and infrastructure
- Regularly update model with new data
- Collaborate with local organizations for support

THANK YOU

We appreciate your attention and welcome any questions or feedback about our project.