# Predicting Water Well Functionality in Tanzania

- Objective: Develop a classification model to determine whether a water well is functional, non-functional, or in need of repair.

- Importance: Helps in water resource management and decision-making for well maintenance.

Business Problem

- Challenge: Many communities in Tanzania rely on wells for water, but some become non-functional.

- Impact: Identifying failing wells in advance can optimize maintenance and resource allocation.

- Solution: A machine learning classification model to predict well functionality based on historical data.

Distribution of Well Functionality
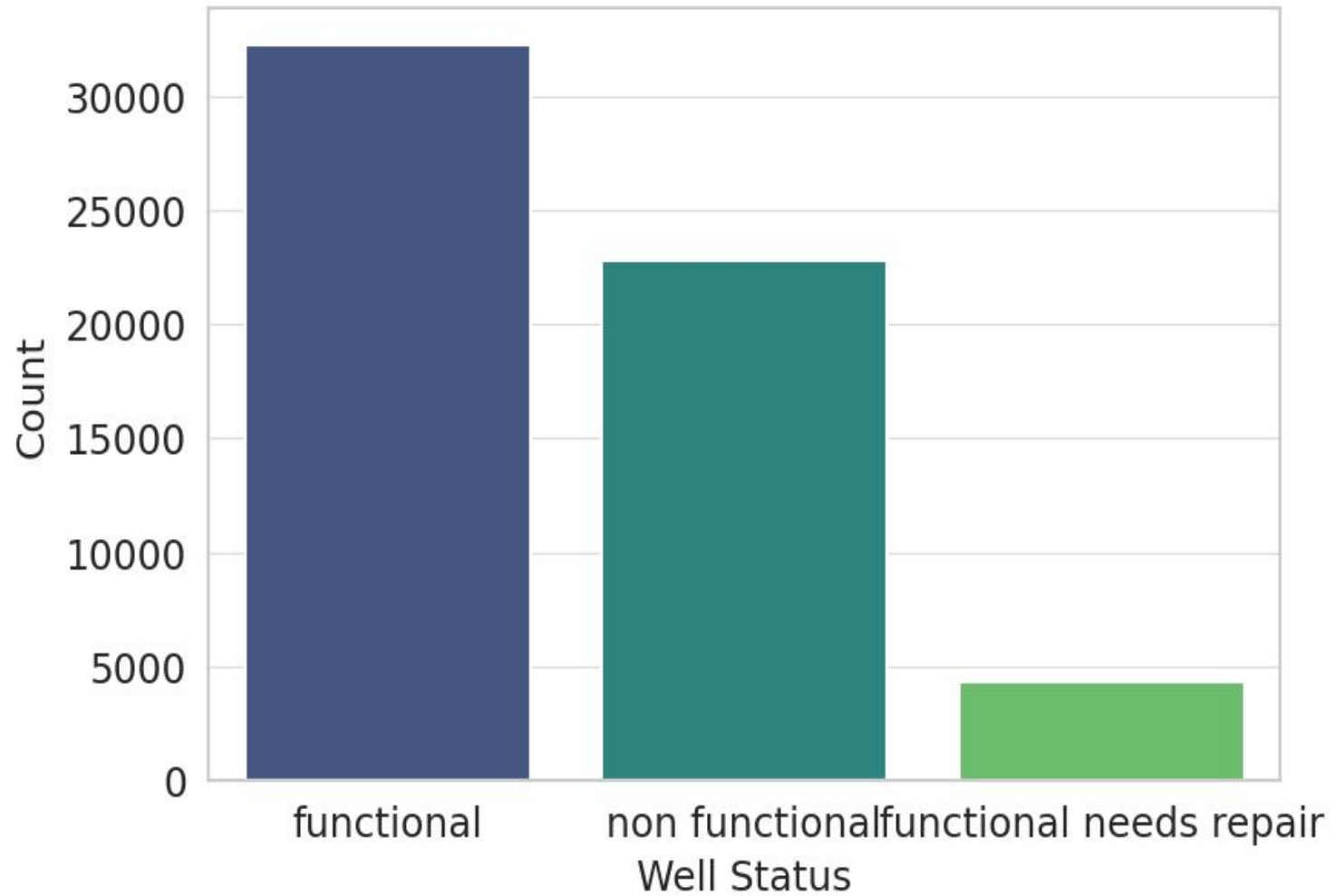
Dataset Overview

- Source: Tanzania water well dataset (59,400 records, 40 features).

- Target Variable: status_group (Functional, Non-Functional, Functional but needs repair).

- Feature Types:

- - Numerical: Population, GPS height, construction year, amount of water (TSH), etc.

- - Categorical: Water source, extraction type, management type, payment type, etc.

Proportion of Missing Data per Feature
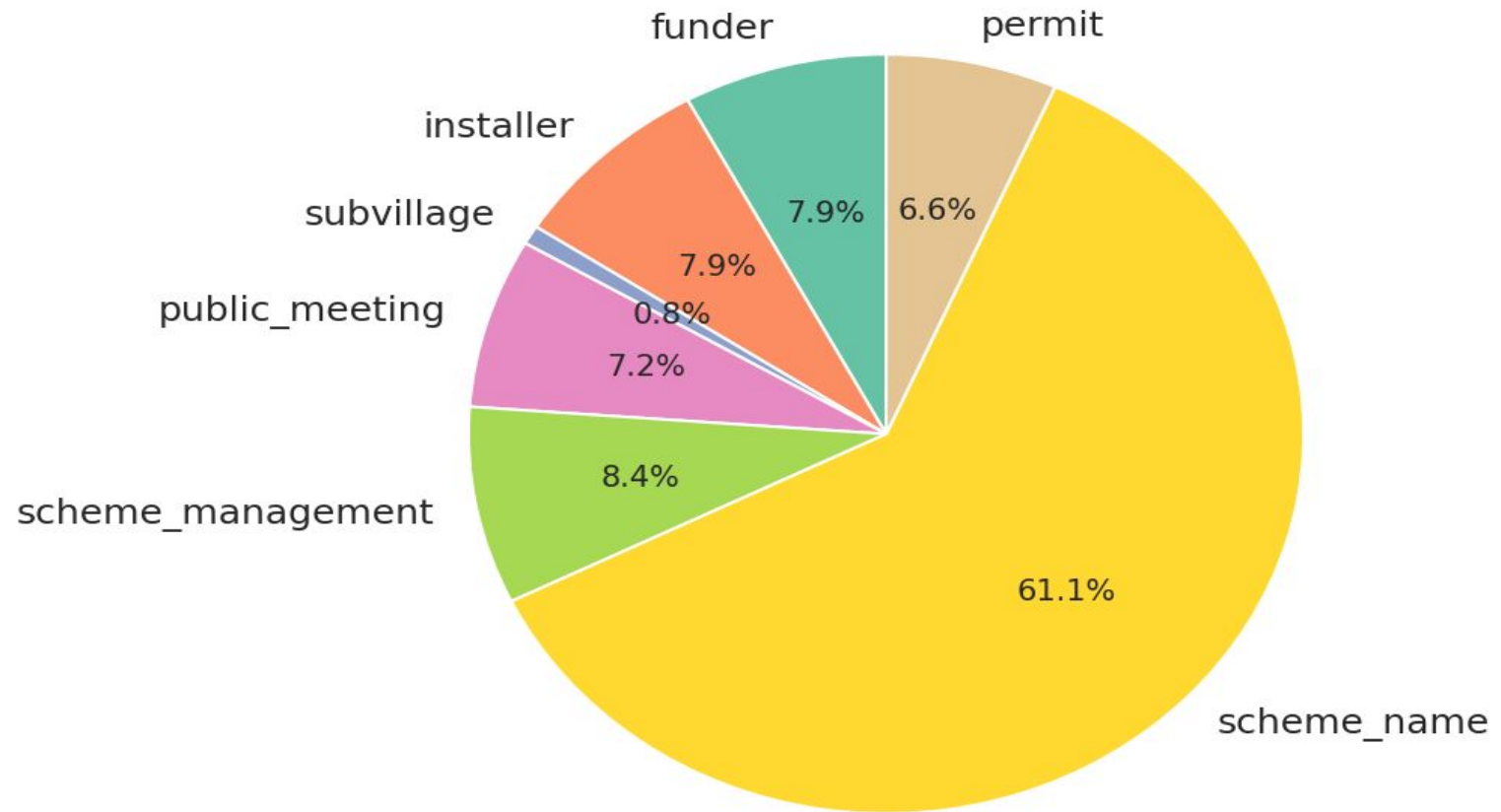
Data Processing

- Handling Missing Values:

- - Columns with significant missing values were reviewed.

- - Imputation strategies applied where necessary.

- Feature Selection:

- - Removal of redundant or irrelevant features.

- - Encoding categorical variables for modeling.

Model Selection & Training

- Models Considered:

- - Decision Tree

- - Random Forest

- - Logistic Regression

- - Gradient Boosting

- Evaluation Metrics:

- - Accuracy

- - Precision & Recall

- - F1-Score

Model Performance Comparison

Key Findings & Results

- Best Model: Random Forest

- Performance Metrics:
- - Accuracy: 72.39%
- - Precision: 71.52%
- - Recall: 72.39%

- Insights:

  The Random Forest model achieved 72.39% accuracy, effectively predicting well functionality but struggling slightly with "needs repair" cases. Key factors like construction year and extraction type were most influential, and further tuning could improve overall precision and recall.

Feature Importance in Prediction
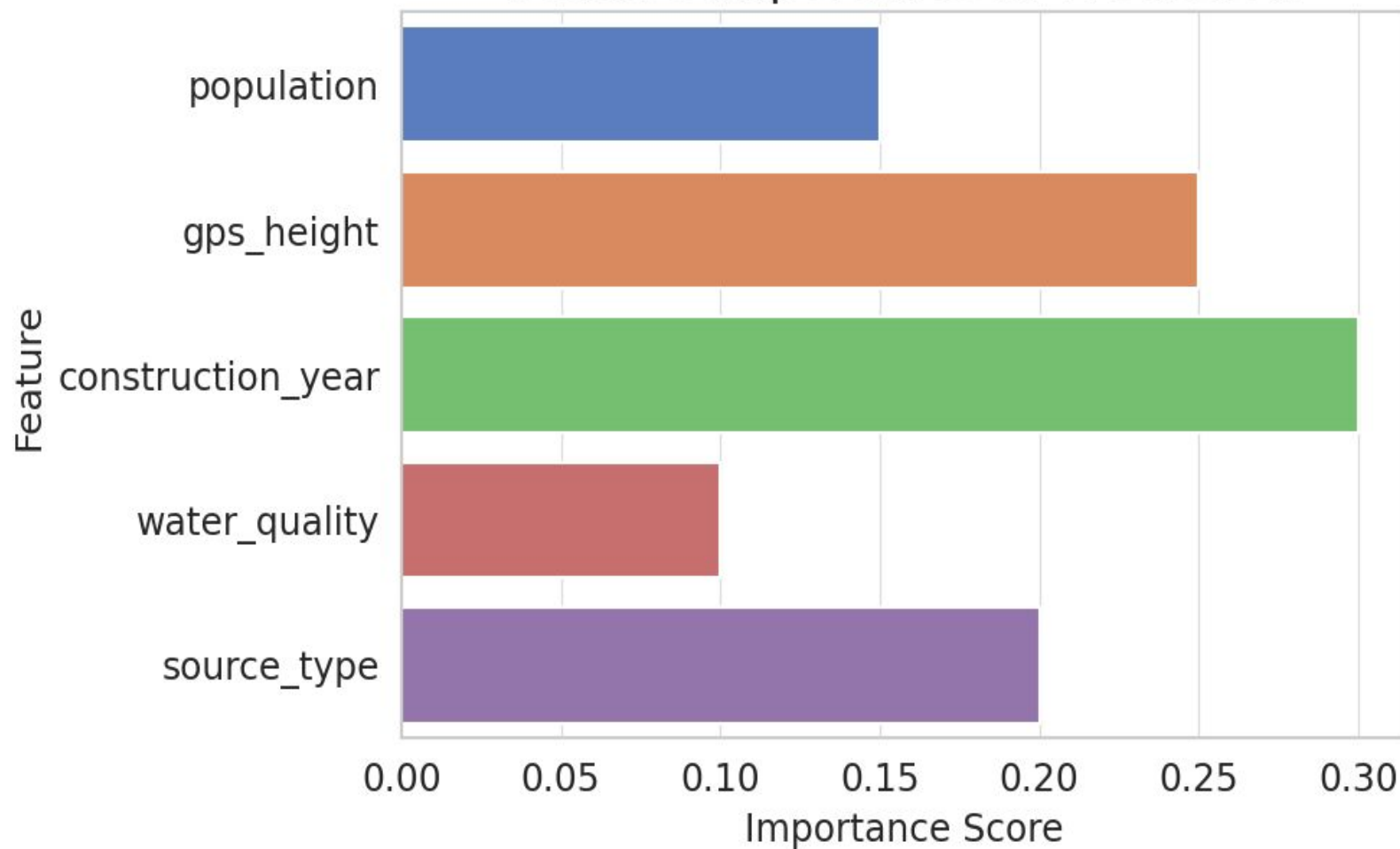
Business Impact

- Efficient Resource Allocation: Prioritizing non-functional wells for maintenance.

- Improved Water Access: Ensuring communities have access to functional wells.

- Data-Driven Decision Making: Supporting policy-making and infrastructure development.

Conclusion & Next Steps

- Conclusion:

- - The model successfully predicts well functionality.

- - Key factors influencing well functionality identified.

- Next Steps:

- - Further fine-tuning of the model.

- - Deployment of predictions for real-time decision-making.

- - Collaboration with government agencies for implementation.