Ujamaa Springs.

Pionering Tanzanias water accessibility through machine learning.

Overview

By delving into the status of wells, this project seeks to contribute data-driven insights to the overarching goal of ensuring water security for the Tanzanian population. Through this innovative approach, Ujamaa Springs endeavors to play a pivotal role in advancing the mission for improved water accessibility and sustainability in the country.

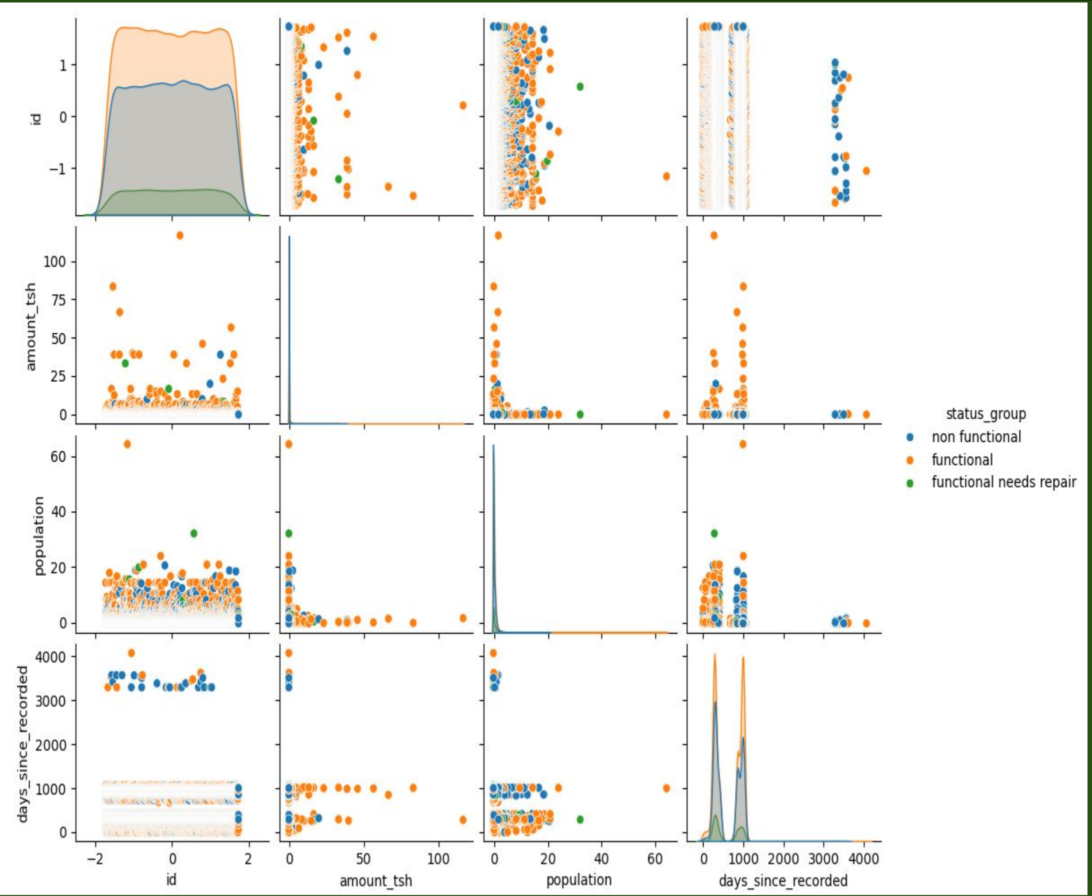
1. Problem statement

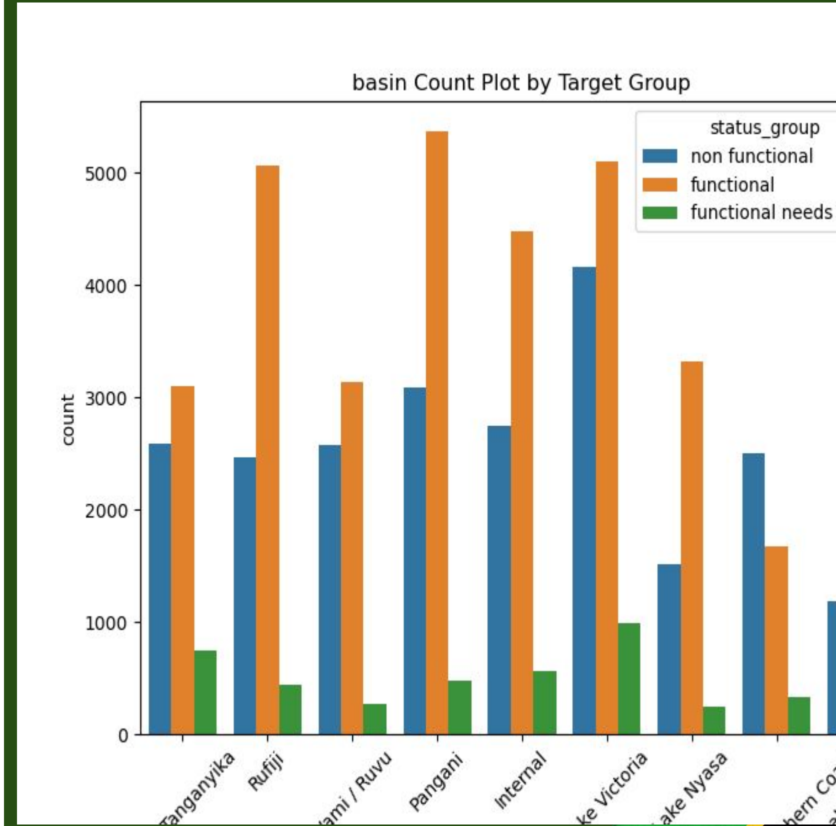
- - Tanzanian communities face a challenge: not enough reliable wells for clean water.
- - The existing wells are not living up to expectations, making it tough for people to get the water they need.
- - It's not just about having more wells; it's about making sure each one works smoothly, like turning on a tap.
- - This problem is causing difficulties for the community, straining their access to clean water.

Objectives

1. Exploratory Data Analysis (EDA): What Factors Influence Well Functionality?
2. Baseline Model - Multinomial Logistic Regression: Can We Predict Well Functionality Probability?
3. Main Model - Random Forest: How Can We Improve Well Functionality Prediction?
4. Recommendations for Model Enhancement: Optimal Features for Improved Well Functionality Prediction

1. EDA - What factors affect well functionality





2. Baseline Model - Multinomial Logistic Regression: Can We Predict Well Functionality Probability?

- Overall Accuracy: 62.85%
- The model correctly predicts well functionality about 62.85% of the time.
- Functional Wells:- Precision: 0.63 ,Recall: 0.81,F1-Score: 0.71
- Functional Needs Repair Wells: Precision: 0.00, Recall: 0.00
- Model struggles due to class imbalance.
- Non-Functional Wells: -Precision: 0.63, Recall: 0.49,F1-Score: 0.55
- The model performs reasonably well in predicting functional and non-functional wells.
- Struggles with wells needing repair, likely due to class imbalance.
- Further optimization, addressing class imbalance, or exploring more complex models is recommended for improvement.

3. Main Model - Random Forest: How Can We Improve Well Functionality Prediction?

- Precision, Recall, F1-Score Breakdown:
  - *Functional Wells:* -Precision: 0.77,Recall: 0.86,F1-Score: 0.81.
  - High precision and recall in identifying functional wells.
  - *Functional Needs Repair Wells:*Precision: 0.49,Recall: 0.23,F1-Score: 0.31
  - Improvement, but challenges persist due to class imbalance.
  - *Non-Functional Wells:* Precision: 0.79,Recall: 0.74,F1-Score: 0.76
  - Maintains good performance in identifying non-functional wells.
- Macro and Weighted Averages:
  - *Macro Average F1-Score:* 0.63 -Slight improvement over Logistic Regression.
  - *Weighted Average F1-Score:* 0.76 -indicates good overall predictive performance.
- Random Forest outperforms Logistic Regression, particularly in identifying functional wells.
- Challenges persist in predicting wells needing repair, possibly due to class imbalance.

4. Model Enhancement: Optimal features for improved well functionality prediction.

1. Feature Engineering:- Explore new features to capture well functionality complexities and consider interactions for more meaningful insights.

2. Imbalanced Data Handling:*Apply advanced techniques (e.g., oversampling, undersampling, SMOTE) to address class imbalance, especially for "Functional Needs Repair."

3. Ensemble Methods:*Investigate stacking and hyperparameter tuning for Random Forest and Gradient Boosting to improve predictive performance.

4. Model Interpretability: Enhance interpretability with SHAP values or feature importance plots for informed decision-making.

5. Cross-Validation Strategies: Experiment with diverse cross-validation methods (e.g., stratified, nested) for robust model performance assessment.

6. Threshold Adjustment: Optimize classification thresholds to balance false positives and false negatives based on project priorities.

Conclusion.

In conclusion, the analysis lays the foundation for informed decision-making in addressing Tanzania's water infrastructure challenges. Ongoing optimization and a proactive approach to model refinement are essential for ensuring accurate and robust predictions, contributing to sustainable improvements in well functionality access.

Thank you.