Henry-Njoroge / **project_tanzania_water_wells**

Code Issues Pull requests Actions Projects Security Insights Settings

Phase 3 Project

1 star 0 forks 0 watching Branches Activity
Tags

Public repository

1 Branch 0 Tags Go to file t Go to file Add file Code

Henry-Njoroge ppt bb7bc86 · 1 minute ago

Datasets	final	9 minutes ago
Notebooks	final	9 minutes ago
Submission folder	ppt	1 minute ago
README.md	Update README.md	1 hour ago

README

Predicting Water Well Functionality in Tanzania

Overview

Access to clean and reliable water remains a pressing development challenge in Tanzania. While thousands of water points have been established, many are non-functional or in need of repair, limiting safe water access for surrounding populations. The goal of this project is to predict the functionality status of water wells—whether functional, functional but needs repair, or non-functional—using available geospatial, demographic, and technical attributes of each well. Accurate predictions can help governments, NGOs, and donors prioritize resource allocation, reducing costs and improving service delivery.

Data Understanding

The dataset comes from the Tanzanian Ministry of Water and includes information on over 59,000 water points across the country. The main files used are:

- Training_set_values.csv: Features for each water point (39 columns).

- Training_set_labels.csv: Target variable (status_group) for each water point. The dataset contains features such as:
- Geographical data: longitude, latitude, region, basin, population served.
- Well characteristics: construction year, extraction type, water source, water quality, management type.
- Administrative data: funder, installer, permit status, scheme management.
- Target variable: water well condition (functional, functional needs repair, non-functional).

Stakeholders

The primary stakeholders for this project are:

1. Government Agencies & Local Water Authorities: To identify regions with high concentrations of non-functional wells for targeted maintenance and repair campaigns, optimizing limited budgets and manpower.
2. Non-Governmental Organizations (NGOs) & Donors: To make data-driven decisions on where to fund new well construction or rehabilitation projects, ensuring resources are allocated to the areas of greatest need and highest potential impact.
3. Community Health Teams: To anticipate water access crises and proactively address them, thereby improving public health outcomes.

Modeling

We developed the following models: Logistics;DecisionTree,Random Forest and XGBoost. The model's performance was evaluated using the F1 score (macro-averaged) as the primary metric. This is because the target classes are imbalanced, and the F1 score provides a balanced measure between Precision (minimizing false positives, e.g., incorrectly labeling a functional well as needing repair) and Recall (minimizing false negatives, e.g., failing to identify a non-functional well). A high recall is particularly important to ensure that few faulty wells are missed during inspection prioritization.

Conclusion

Prioritize Older Wells: Wells constructed before the 1990s are at a higher risk of failure and should be prioritized for inspection.

Focus on Specific Management Models: Wells under specific management groups (e.g., 'user-group' vs. 'commercial') may require different support structures.

Target Regions with High Failure Rates: Direct resources to geographical basins with a historically high proportion of non-functional wells.

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

- **Jupyter Notebook** 100.0%