Capstone Two - Project Proposal - Water Pump Status in Tanzania

Problem Statement:

How can the government of Tanzania improve water pump maintenance by knowing the pump functional status in advance?

Context:

Tanzania is a developing country and access to water is very important for the health of the population. For this reason, it is vital that all water pumps are properly working. Currently, the only way to monitor pump working status is by physically visiting the site. This is time consuming and costly. Therefore, a more intelligent solution to monitor water pump status is desirable.

Criteria for Success:

Given water pump data, the Ministry of Water for the government of Tanzania will be able to know the water pump status in advance. The final models will be evaluated using the classification metrics discussed below.

Scope of solution space:

The models to predict water pump status will be developed using crowd sourced data provided by Taarifa. Taarifa is an open source platform for reporting and triaging infrastructure issues.

Constraints:

Since this analysis relies on crowd sourced data, it depends on citizens actively providing information to Taarifa and that that information is accurate.

Stakeholders:

- Minister of Water, Hon. Jumaa H. Aweso
    - This is the most senior position in the government related to water issues. Therefore, he should be aware of improvements to water pump maintenance.
- Deputy Minister for Water, Hon. Maryprisca Mahundi (Mp)

    - This person works directly under the Minister of Water and should also be made aware of water pump maintenance improvements.
- Permanent Secretary, Eng. Anthony Sanga
    - The Permanent Secretary is in charge of the water management team. Since this person works more directly with individuals managing water issues, they should also be involved in any decisions related to water pump maintenance.

Data Sources:

The data is provided by Taarifa. The dataset consists of 59400 rows and 39 features. Data types include strings, numerical, datetime, and categorical. The outcome variable, water pump status, consists of three categories: functional, functional needs repair, and non-functional.

Modeling Approach:

Since the outcome variable for this problem consists of three classes, I will use multiclass classification models to solve the business problem. During the data analysis stage, I will determine the feature importance as it relates to water pump status. Evaluation metrics will include accuracy, precision, recall, and F1 score. A range in different model algorithms will be tried. An exploration in train-test split sizes and number of cross validation folds will be done to find the best model for each algorithm. A final set of models will be chosen depending on their evaluation metrics.

Deliverables:

- At the end of this project, the client will be presented with the following deliverables:
    - An analysis of water pump status as it relates to water pump features.
    - A description of which features are important in predicting the outcome variable (in this case, water pump status).
    - A measure of feature impact as it is related to predicting water pump status.
- The project itself will include:
    - All Jupyter notebooks used for carrying out the project.
    - A slide deck presentation of the project results.
    - A final project report.