ASSESSMENT OF THE CONDITIONS OF WELLS IN TANZANIA

EE. NGO

Introduction

Tanzania, a developing country, has a problem in providing water to its fast-growing population of

57 million. The country has already established wells that are expected to provide the much-needed

water, however, some of the wells require repairing for the goal of enough water supply to be met.

Enthusiastic Environmentalists (EE), a renowned NGO is interested in locating the wells that require

repairing and repair them, and enable the country to curb the water problem

I am the data scientist
 tasked with developing a predictive model to know the wells requiring repair. This will enable
 them to cut the cost of surveying and pinpoint the exact wells that will require any repair as using the
 model makes it easier to identify the wells without having to spend on nationwide physical assessment.

Data

The data used in creating the model was of the wells having the condition of the well as the target variable and the rest is the different information about the wells. The target variable had to be converted to a binary target variable from ternary, to have the status of the well as either one that needs to be repaired or one that does not need repair. The next step is to carry out feature engineering where all the columns with null values were dropped. I then preprocessed the data, first onehotencoding the categorical data and then scaling the numeric values.

Modeling and Evaluation

After processing the data and getting it ready for modeling, we head into creating predictive models from which the model with the best performance will be selected and used by the NGO in making predictions. The best metric for this model was the AUC because the target variable is a binary variable. The baseline model was a logistic regression and the score of the model was 77.8%. An improved model was a decision tree which had a score of 78.4%.

Conclusion and Recommendation

The NGO should use the improved model (The Decision Tree Classification model) to carry out the process of identifying the wells that require repairing. This will enable the NGO to save on the cost of carrying out a survey. This also saves time as it is simple to pinpoint the wells instantly with 78% assurance.