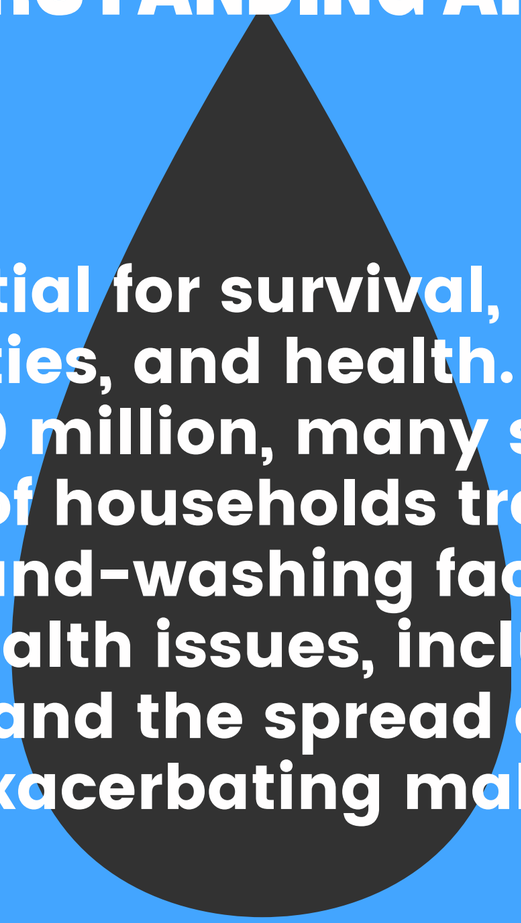
TANZANIA'S OPERATIONAL WATER WELLS PREDICTION

BUSINESS UNDERSTANDING AND OVERVIEW



Water bodies are essential for survival, supporting household use, economic activities, and health. In Tanzania, with a population of around 60 million, many still struggle to access clean water. Only 30.6% of households treat water properly, and 22.8% have adequate hand-washing facilities. Poor sanitation results in significant health issues, including 432,000 annual deaths from diarrhea and the spread of neglected tropical diseases, exacerbating malnutrition.

PROBLEM STATEMENT

OUR NGO, DANIDA, IS FOCUSED ON FINDING AND REPLACING WATER WELLS THAT NEED REPAIR. OUR CLASSIFICATION MODEL IS BEING USED TO BETTER PREDICT WHICH WELLS IN THE AREA ARE OPERATIONAL, NEED REPAIRS OR ARE NON-FUNCTIONAL USING VARIOUS INFORMATION SUCH AS WHEN EACH WELL WAS INSTALLED, WHO FUNDED THE PROJECT, AND POPULATION AROUND EACH WELL. ANY IMPROVEMENT IN DETERMINING THE BEST WELLS TO INSTALL OR IN THE PREDICTABILITY OF WHICH WELLS NEED REPAIRS COULD HAVE AN ENORMOUS IMPACT ON THE PEOPLE OF TANZANIA.

OBJECTIVES

To aid in improving maintenance operations by focusing inspections on the water points that have a high likelihood of requiring repair or having failed altogether

To determine the functionality status concerning payment type

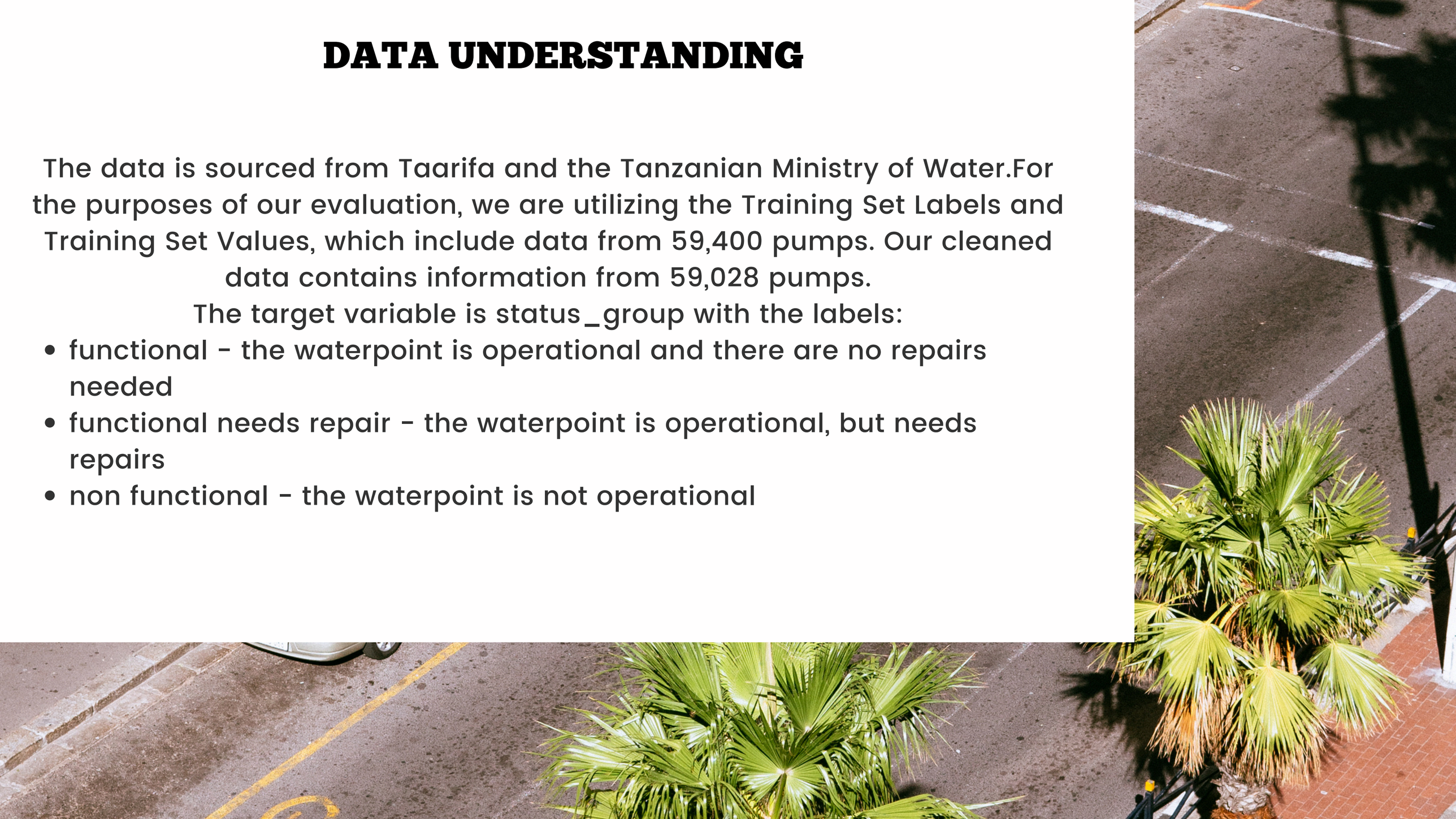
To provide 70%–75% accurate predictions on the functionality of wells

DATA UNDERSTANDING

The data is sourced from Taarifa and the Tanzanian Ministry of Water.For the purposes of our evaluation, we are utilizing the Training Set Labels and Training Set Values, which include data from 59,400 pumps. Our cleaned data contains information from 59,028 pumps.

The target variable is status_group with the labels:

- functional - the waterpoint is operational and there are no repairs needed
- functional needs repair - the waterpoint is operational, but needs repairs
- non functional - the waterpoint is not operational



DATA PREPARATION

During this phase of the analysis, the focus will be on addressing missing values, identifying and removing duplicated entries, resolving inconsistencies, and handling invalid data. Initially, we excluded irrelevant columns identified during the data understanding phase, as there is no need to prepare columns that will not be utilized in the analysis.

EXPLANATORY DATA ANALYSIS

UNIVARIATE ANALYSIS

It helped in understanding the distribution of individual variables (e.g., number of wells that are functional vs. non-functional) and also.It provides insights into the central tendency (mean, median) and variability (range, standard deviation) of the data.

BIVARIATE ANALYSIS

We Explored relationships and correlations between two variables (e.g., the relationship between well functionality and payment type).and also Provides insights into how different factors together influence the target variable, aiding in building more accurate predictive models.

MODELLING

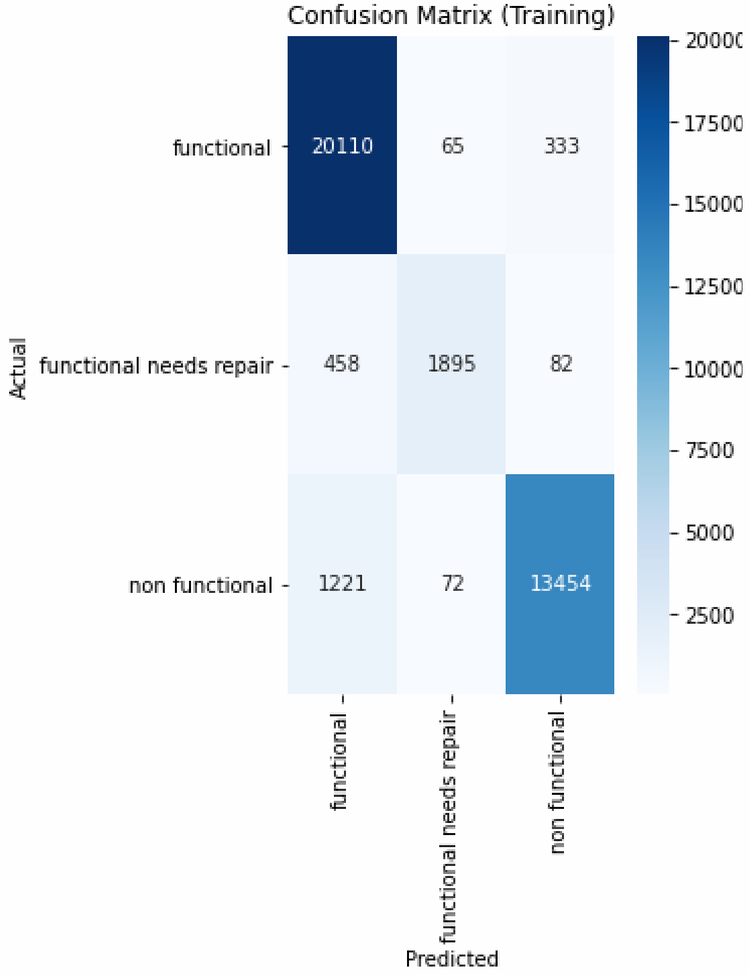
To prepare our data to machine learning, we did some feature engineering, encoding.
The following models were used:
Decision trees
Random forest
XG boost

we split the data the following ways

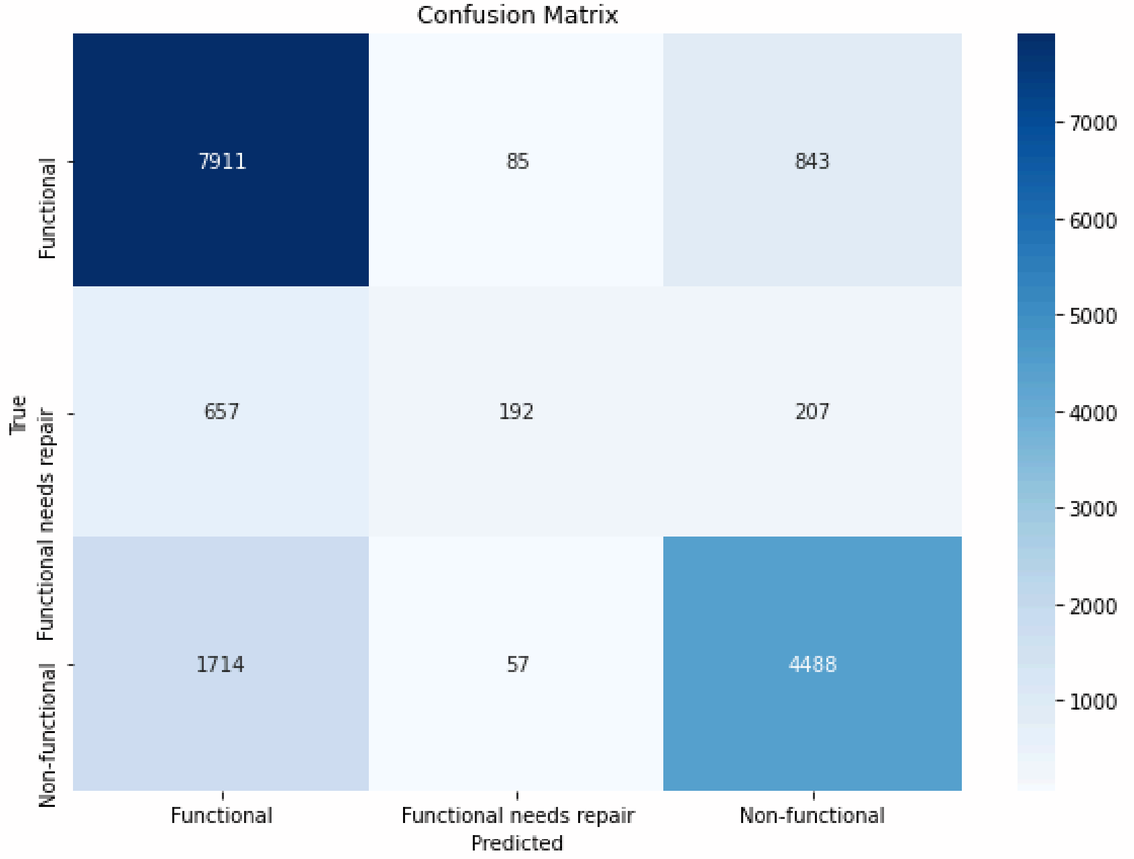
70% Training **30%** Testing

RESULTS OF THE MODELS

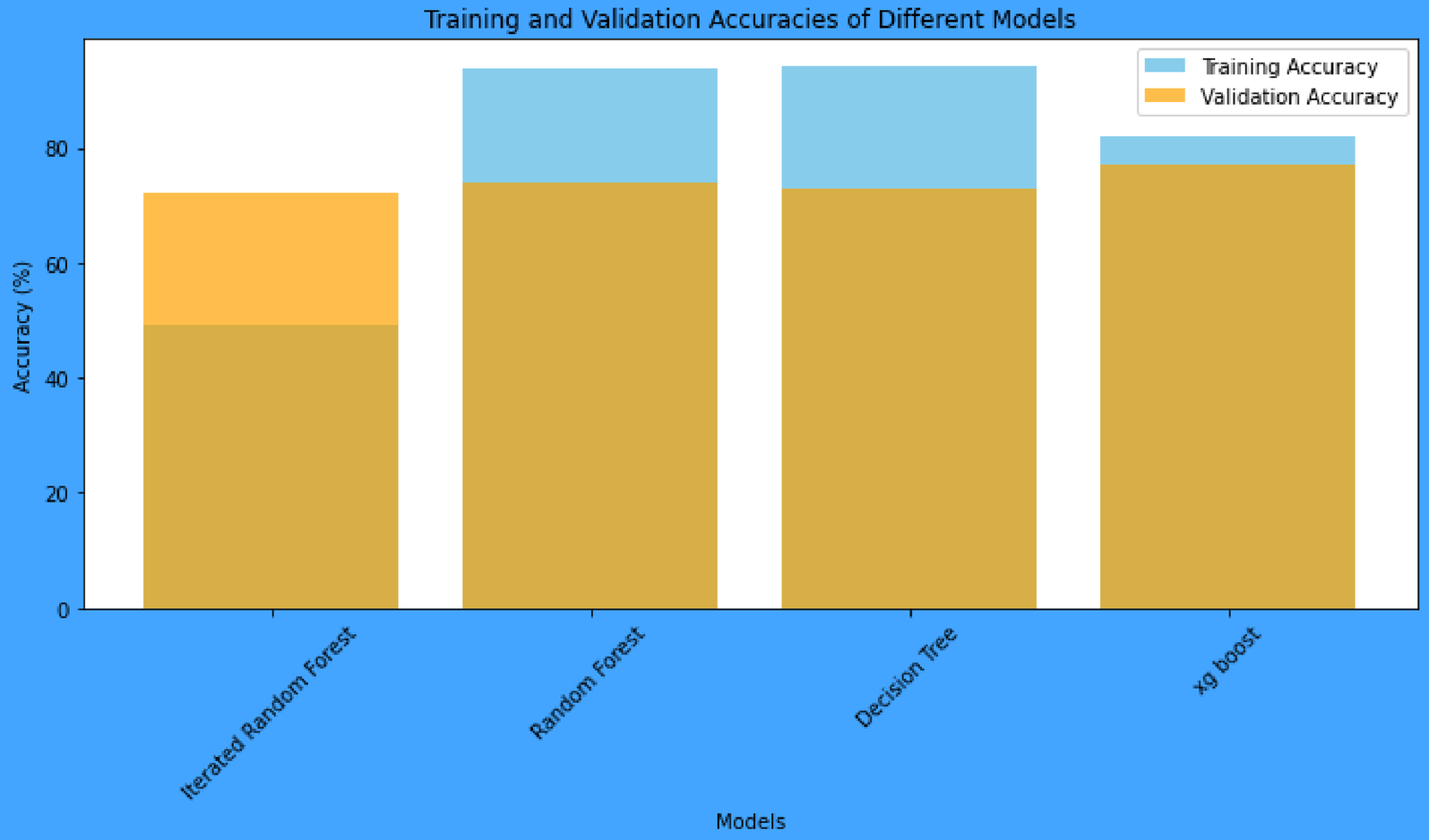
DECISION TREE MODEL



XG BOOST



RESULTS OF THE ALL MODELS COMBINED



XGBoost appears to be the best model overall, as it shows a good balance between training and validation accuracies, indicating that it generalizes well to unseen data.

RECOMMENDATIONS

Investments should be directed towards communal standpipes, rather than communal standpipe multiples, as the majority of the latter are non-functional.

Priority should be given to non-functional wells and functional wells that need repair, provided they have sufficient water.

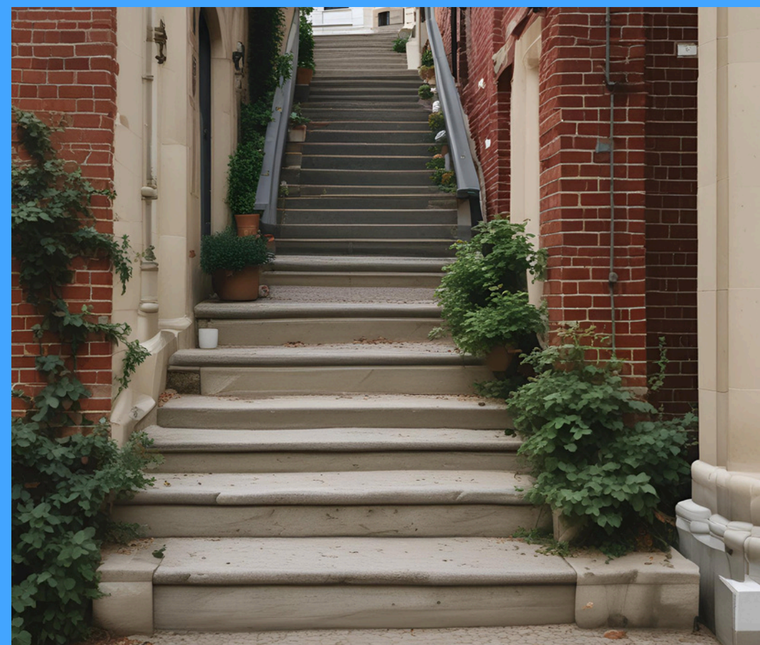
Providing payment creates an incentive and means to maintain wells in a functional state.

NEXT STEPS

Better data trained in our model will improve the predictions

Learn cost of repairs, construction, and preventive maintenance

Create a cost-benefit function to prioritize actions



HELLEN SAMUEL

Thank you!

hellensamuel37@gmail.com