

PREDICTING FAULTY PUMPS

DATA MINING FOR
SUSTAINABLE WATER
MANAGEMENT



MIGUEL BARRIOLA ARRANZ

05/21/2024

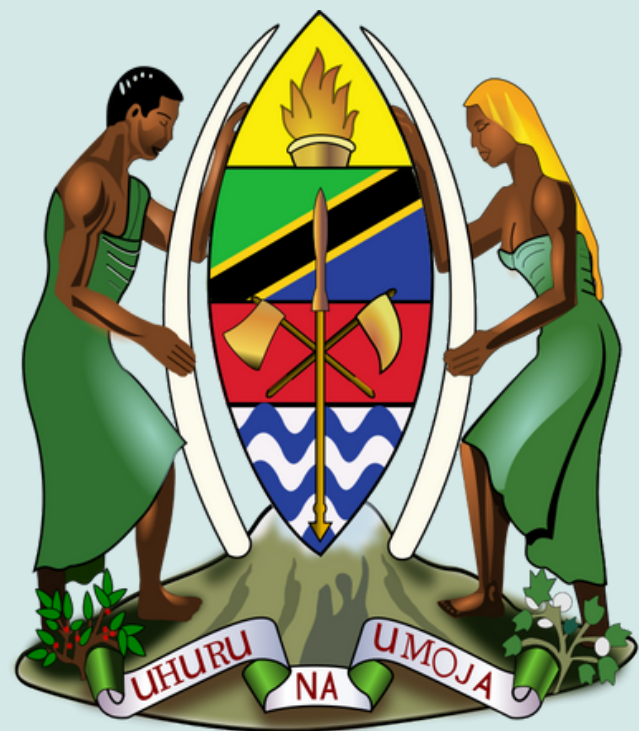
INTRODUCTION

- Project to predict the status of water pumps in Tanzania
- Objectives of the analysis:
 - Classify pumps in Tanzania as functional, or non-functional
 - Improve water access in Tanzania



STAKEHOLDERS

- Stakeholders, including **government agencies and NGOs**, will use these findings to prioritize and streamline efforts towards ensuring reliable water access.
- Primary stakeholders for this project are the **Tanzanian government and international development organizations** focused on improving water access in the region.



BUSINESS CASE

- Core objective:
 1. Enable the identification of functional and non-functional pumps in Tanzania
- Results implications:
 1. Guide decisions on maintenance, investments, and resource allocation.
 2. Support sustainable water management in Tanzania.



DATA

DATASET OVERVIEW

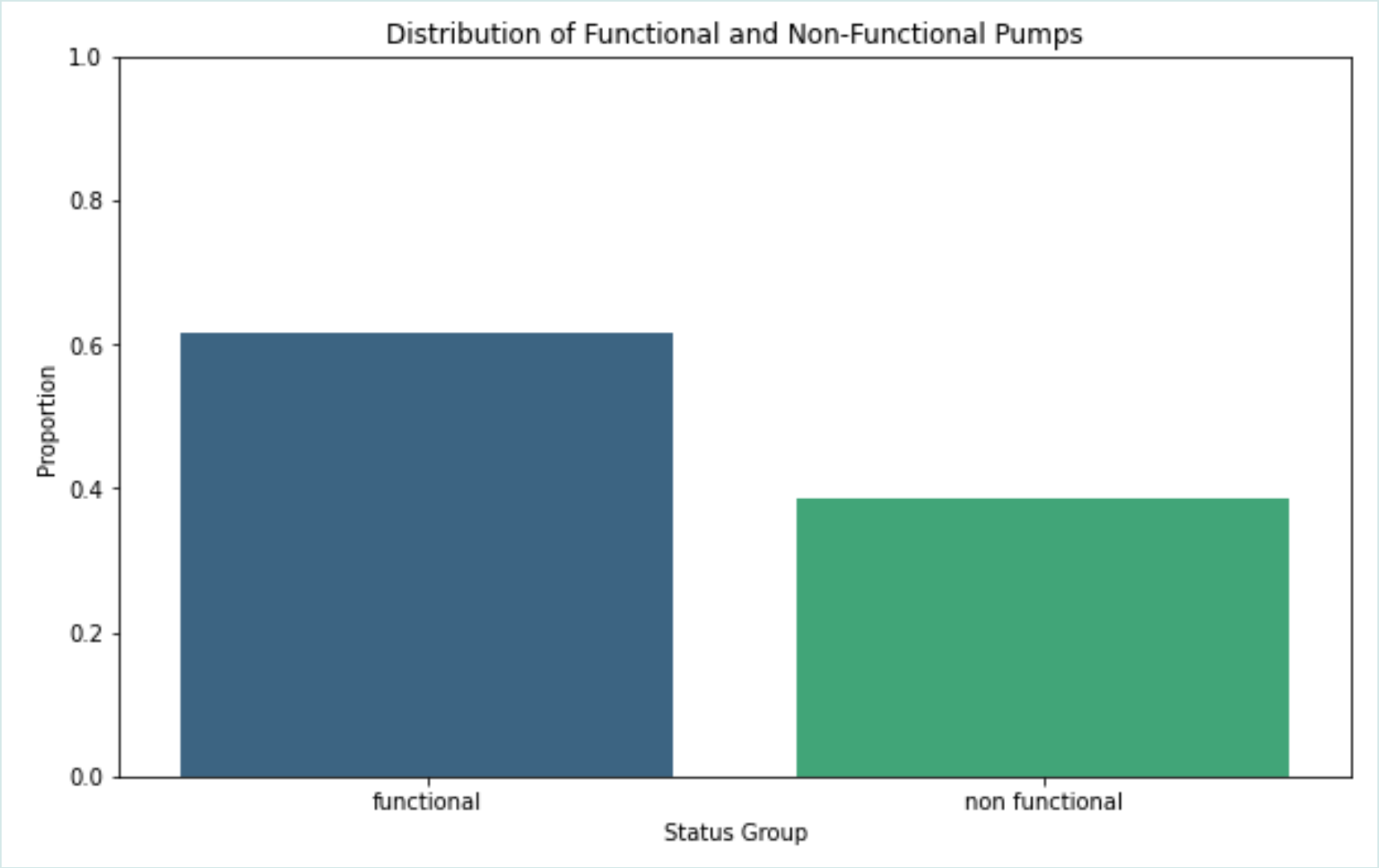
Driven Data provided the following datasets:

- SUBMISSIONFORMAT
- TEST_SET_VALUES
- TRAINING_SET_LABELS
- TRAINING_SET_VALUES

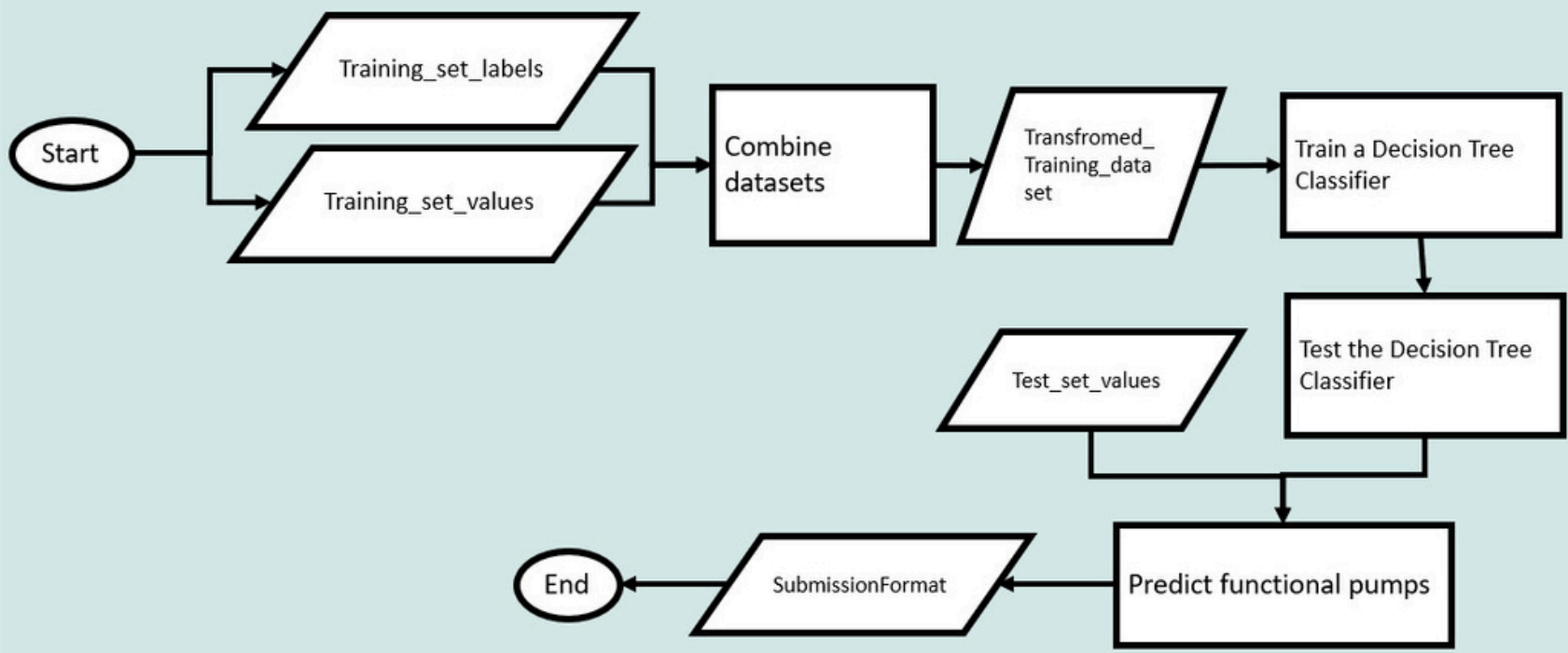
DATASET DESCRIPTION

- TRAINING_SET_LABELS and TRAINING_SET_VALUES were transformed for model building
- Same transformations applied to TEST_SET_VALUES for predictions
- SUBMISSIONFORMAT contains the predicted pump status

PUMPS STATUS

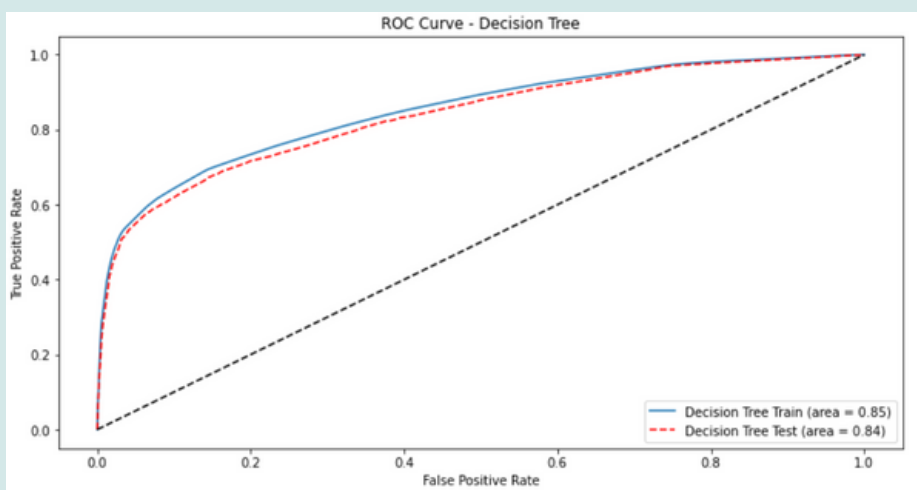


PROJECT OVERVIEW

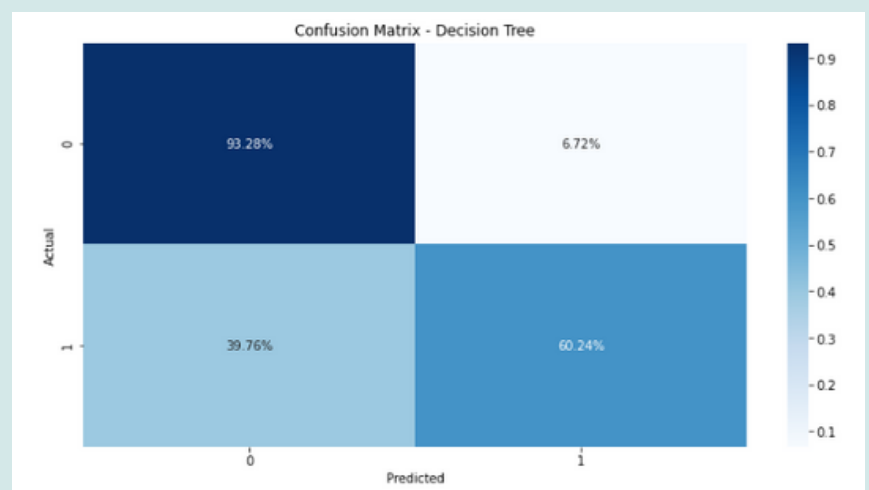


MODELING APPROACH

- Models tested:
 - Logistic Regression
 - Decision Tree
- Evaluation metrics considered were ROC, AUC, and the confusion matrix



Hyperparameter tuning process to get the optimal parameters

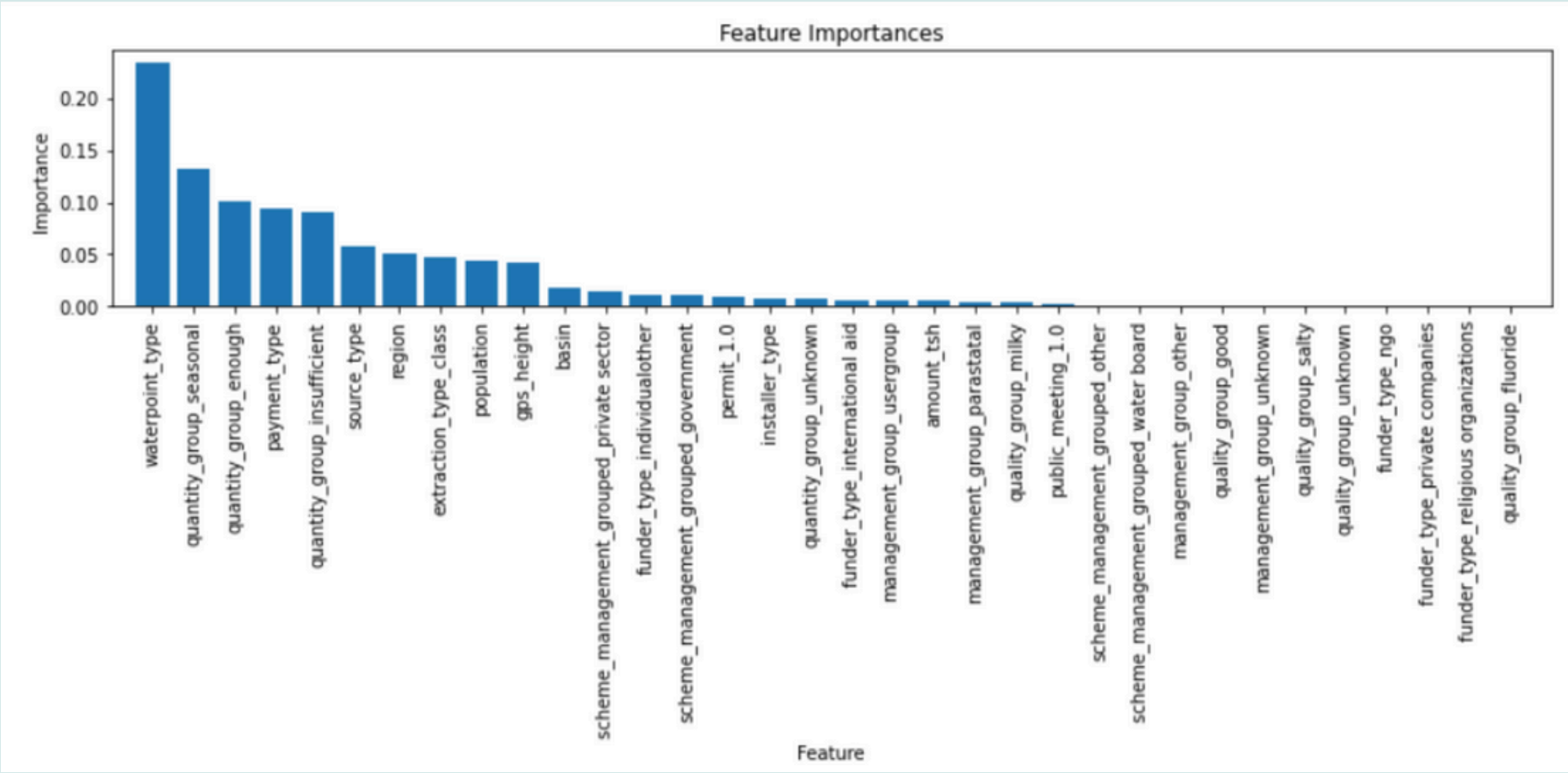


Focus on reducing false negatives for prompt repairs

FEATURE IMPORTANCE

- Here are the most important variables that better discriminate between functional and non-functional:

- a. waterpoint_type
- b. quantity_group
- c. payment_type



RECOMMENDATIONS

- **Align Payment Plans:** Use monthly or per-bucket payment plans like those of functional pumps to increase functionality
- **Use Dry Pumps as Indicators:** Identify non-functional pumps using the presence of dry pumps to focus repair efforts
- **Identify Non-Common Types:** Use pumps without common water point types (e.g. cattle trough) to find and prioritize non-functional pumps for repairs

THANK YOU VERY MUCH

Miguel Barriola Arranz

- LinkedIn:

<https://www.linkedin.com/in/miguel-barriola-arranz/>

- Medium:

<https://medium.com/@mbarriolaarranz>