PREDICTING FAULTY PUMPS

DATA MINING FOR SUSTAINABLE WATER MANAGEMENT

MIGUEL BARRIOLA ARRANZ

05/21/2024

INTRODUCTION

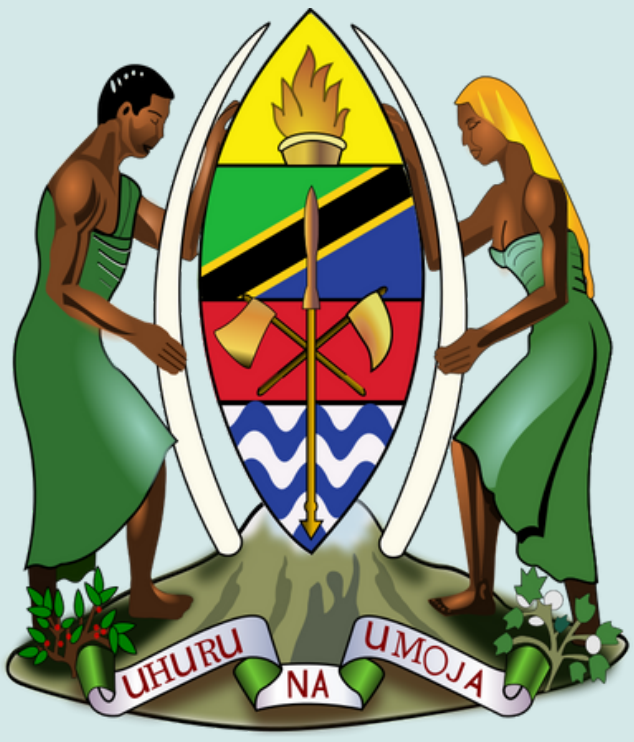
- Project to predict the status of water pumps in Tanzania

- 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

- The core objective is to enable the identification of water pumps in Tanzania that are functional or are non-functional.

- The insights derived from this analysis will directly influence decisions regarding maintenance, investments, and resource allocation in the water infrastructure sector.

- The ultimate goal is to support sustainable water management practices that can significantly impact public health and economic development in Tanzania.

DATA

DATASET OVERVIEW

Driven Data provided the following datasets:

- SUBMISSIONFORMAT
- TEST_SET_VALUES
- TRAINING_SET_LABELS
- TRAINING_SET_VALUES

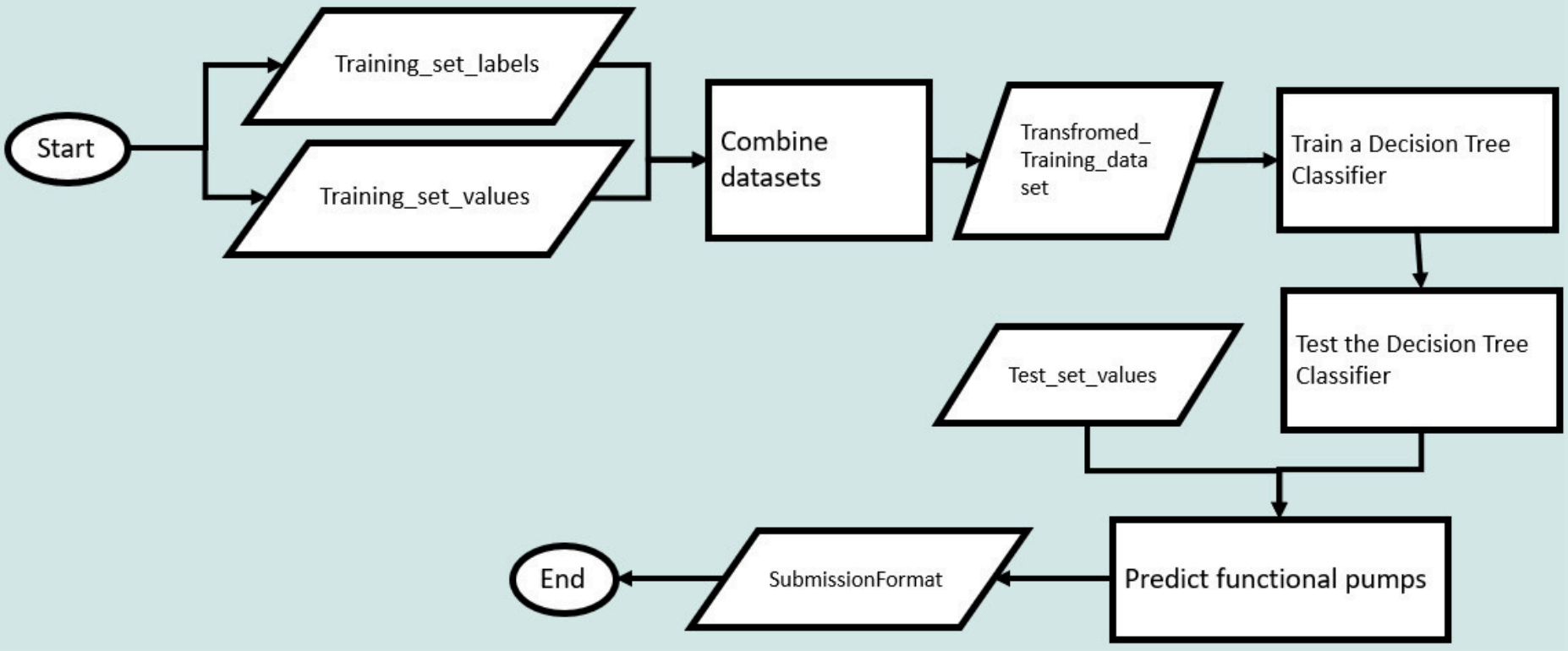
DATASET description

- The TRAINING_SET_LABELS and TRAINING_SET_VALUES were transformed, and models were built based on these transformed values.

- The same transformations were applied to the TEST_SET_VALUES, based on which predictions were made.

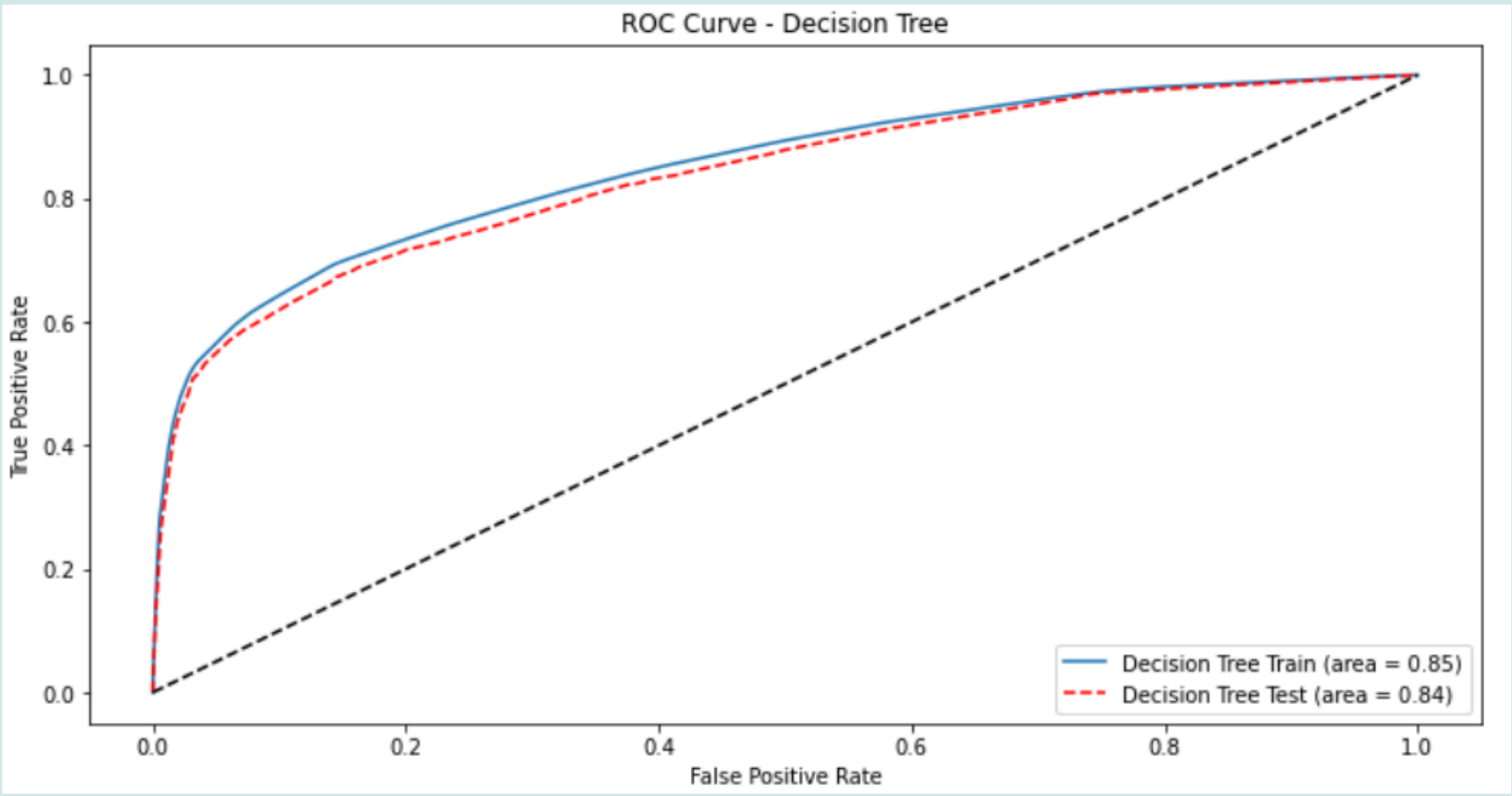
- The SUBMISSIONFORMAT contains the predicted status of the pumps.

PROJECT OVERVIEW



MODELING APPROACH

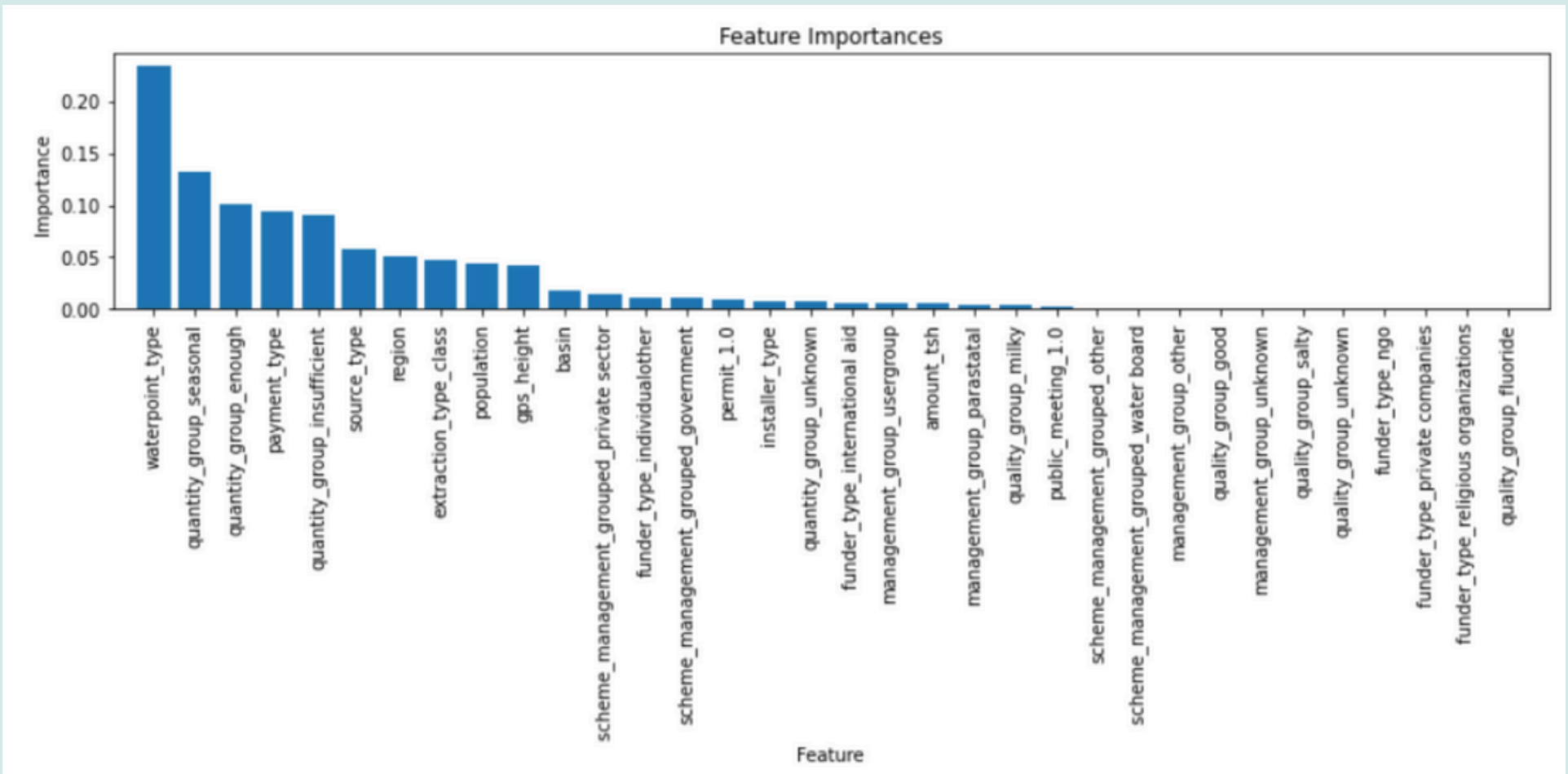
- Models tested:
  - Logistic Regression
  - Decision Tree
- Evaluation metrics considered were ROC and AUC
- Hyperparameter tuning process to get the optimal parameters



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

1. First Recommendation:

Consider aligning the payment plans of water pumps with the more common monthly or per-bucket payment plans used by most functional pumps to increase their chances of being functional.

2. Second Recommendation:

Using the presence of dry pumps as an indicator can help identify non-functional pumps, allowing efforts to be focused on repairing these specific pumps.

3. Third Recommendation:

Identifying pumps that do not have common functional waterpoint types (such as cattle trough, communal standpipe, etc.) can serve as a proxy to determine non-functional pumps and prioritize them for repairs.

THANK YOU VERY MUCH

Miguel Barriola Arranz

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

- Medium: https://medium.com/@mbarriolaarranz