



Firefighting Operational Cost Prediction with Tree-Based Models

Haniff, Ryan
Mahajan, Ravish
Park, Haelang
Ullmann, Gabriel

CONTENTS

1. **Introduction: Research Questions**
2. **Dataset**
 - a. Feature Selection
 - b. Cost Categories
 - c. Data Preprocessing
3. **Model Design**
4. **Results**
 - a. Summary
 - b. F1-Score Comparison
 - c. Classification Comparison by Category
5. **Conclusion and Discussion**
 - a. Comparing with the State of the Art
 - b. Rooms for Improvement

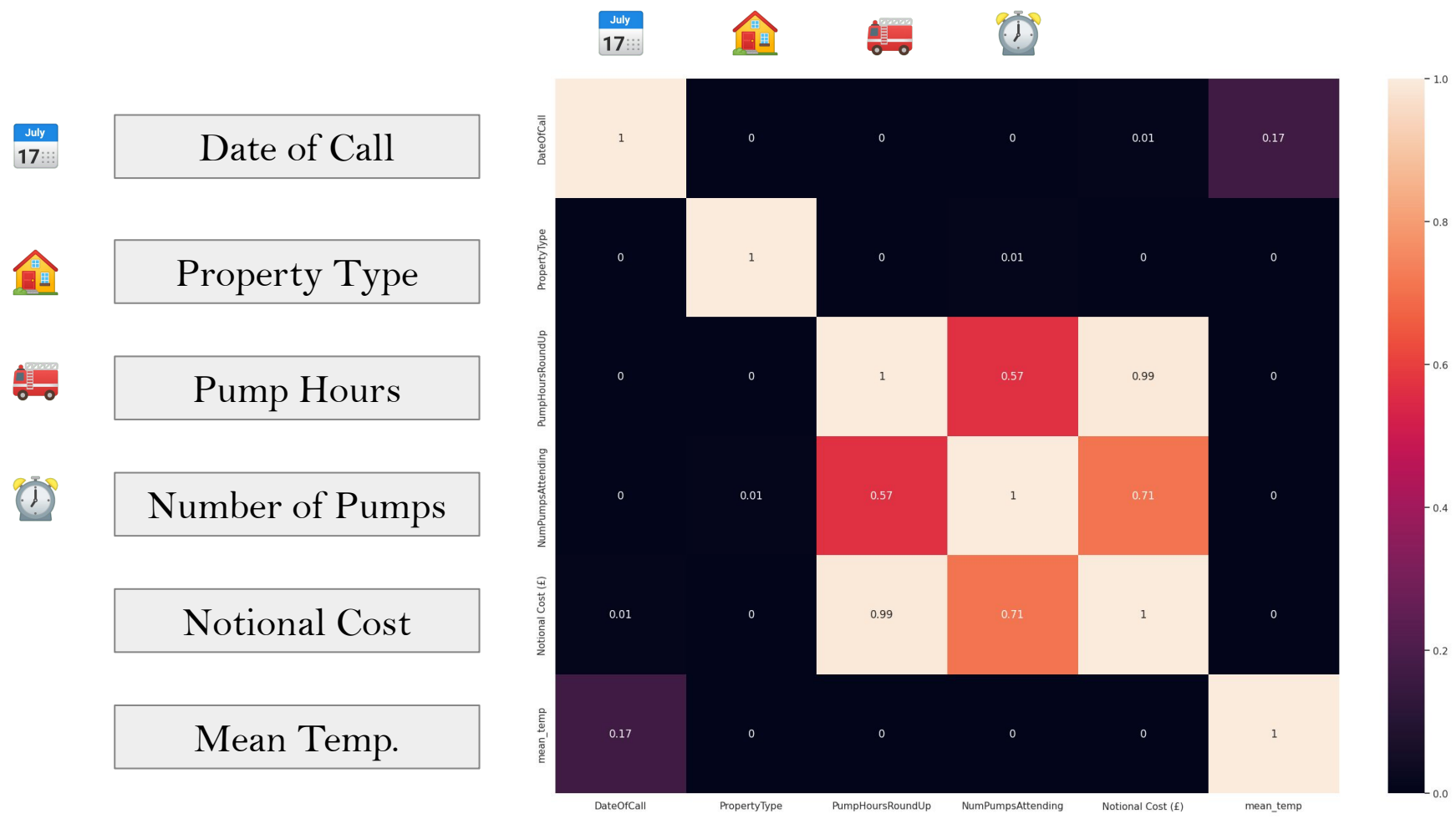


INTRODUCTION

1. Is it possible to **predict the approximate cost of firefighting operations** using tree-based models?
2. **What features correlate more or less** strongly with the operational cost?
3. **Which tree-based model provides the best prediction accuracy?**



FEATURE SELECTION



COST CATEGORIES

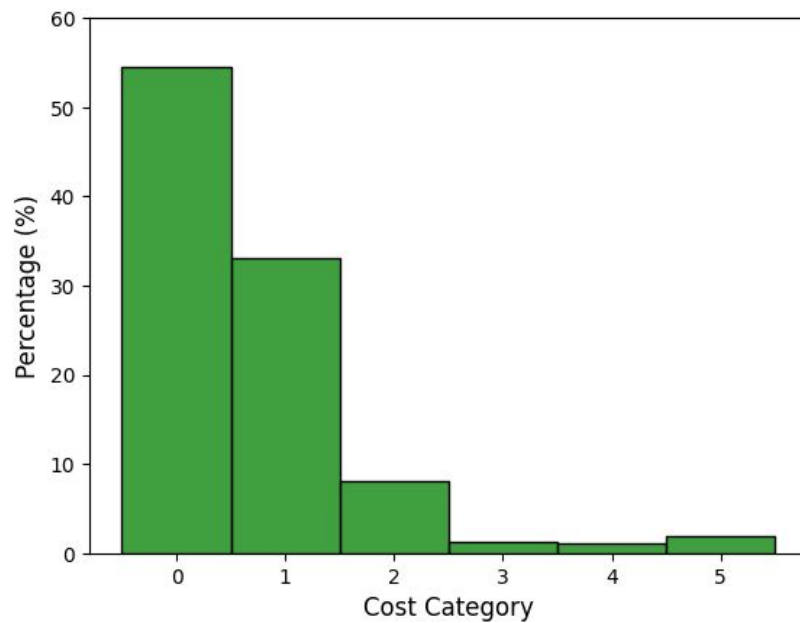
Category	Cost range
0	Cost < £300
1	Cost >= £300 and Cost < £500
2	Cost >= £500 and Cost < £700
3	Cost >= £700 and Cost < £900
4	Cost >= £900 and Cost < £1100
5	Cost >= £1100

DATA PREPROCESSING

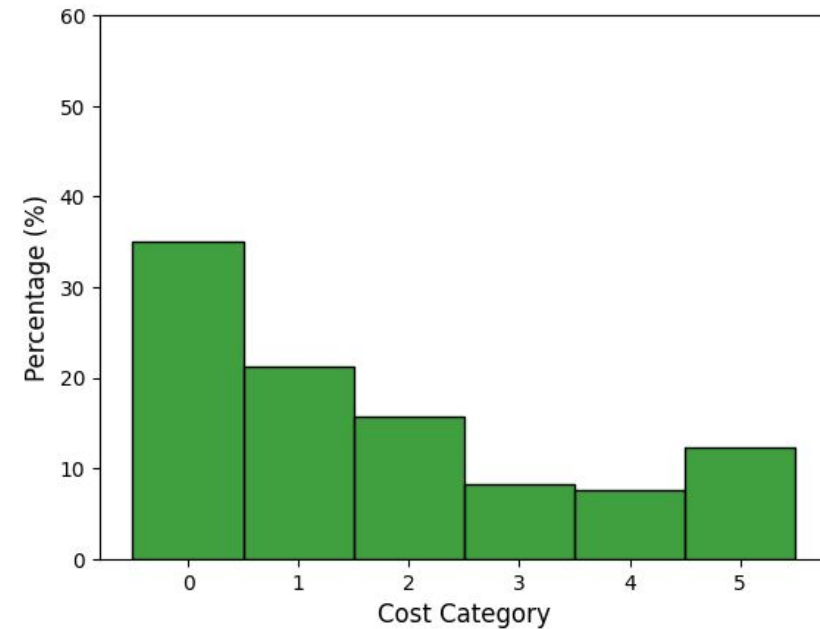
We removed NaN rows, standardized the date formats, and decreased cost category imbalance.

Original Dataset

0	54.49%
1	33.04%
2	8.11%
3	1.27%
4	1.17%
5	1.92%

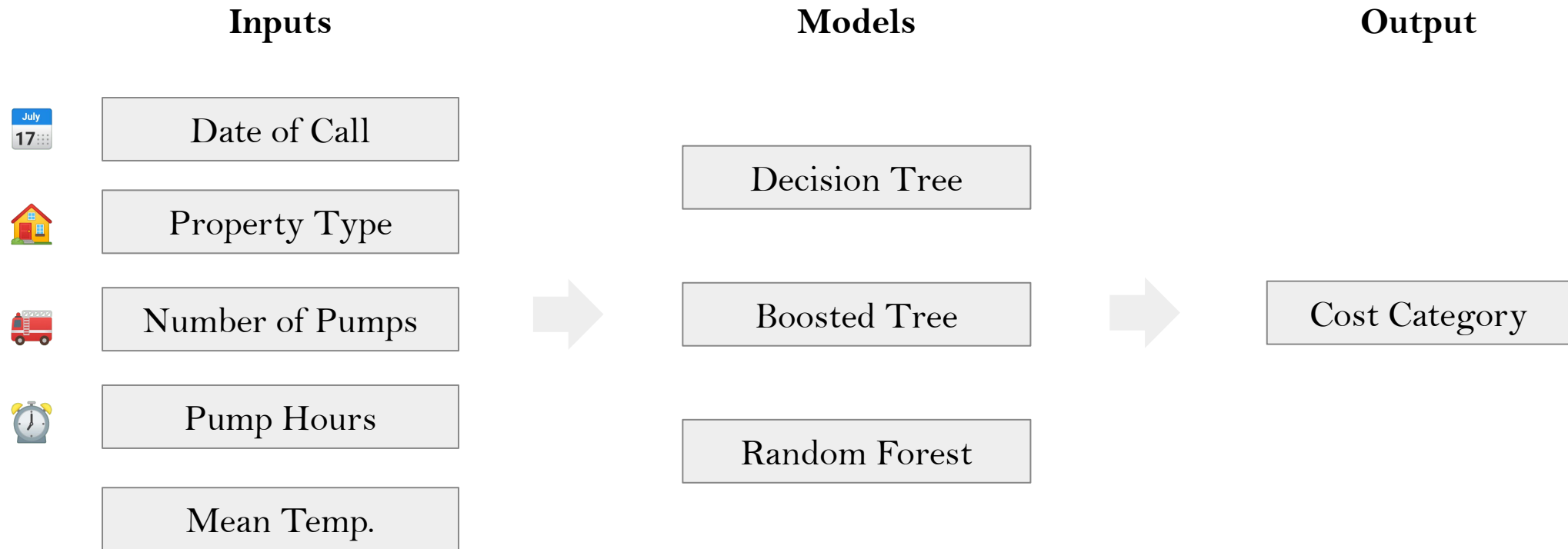


Undersampled Dataset



0	35.05%
1	21.26%
2	15.64%
3	8.19%
4	7.51%
5	12.35%

MODEL DESIGN

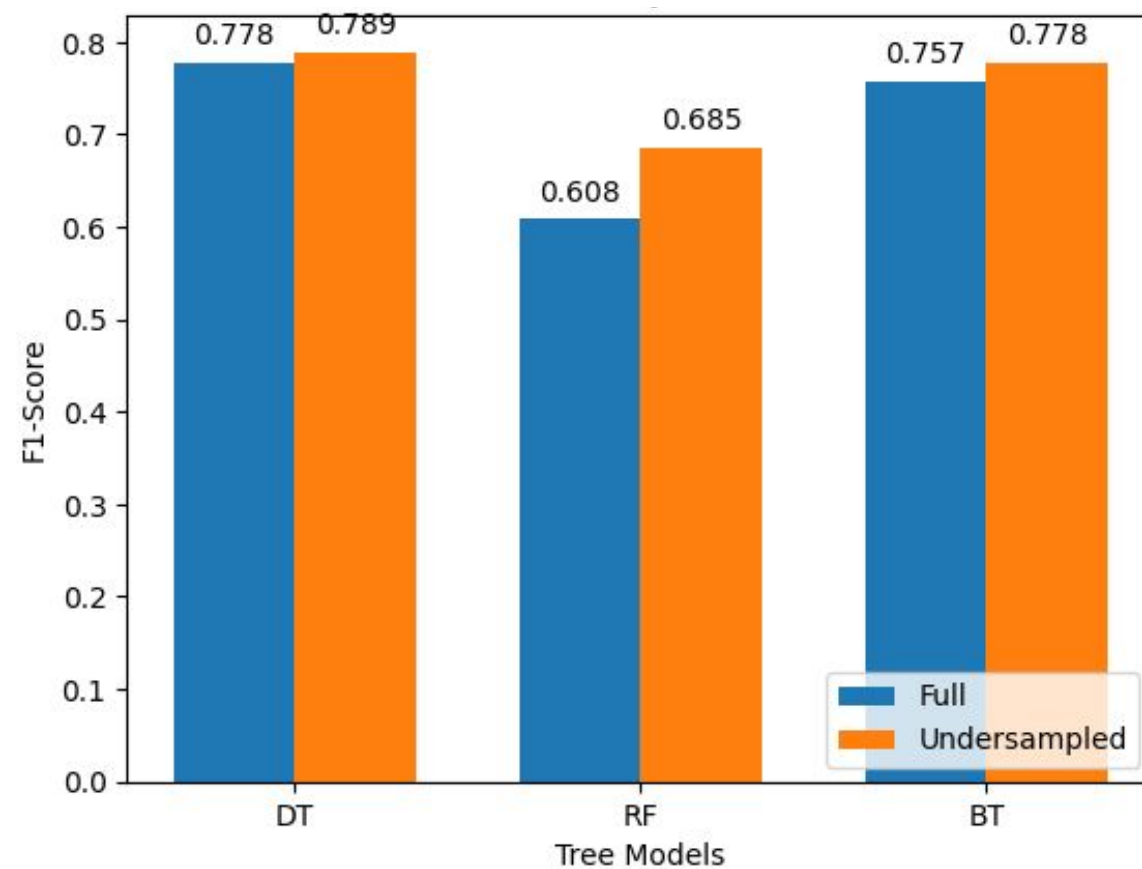


SUMMARY OF RESULTS

After training the models with default parameters and further improving them with hyperparameter search, we observed that the best model is **Decision Tree (0.789)**.

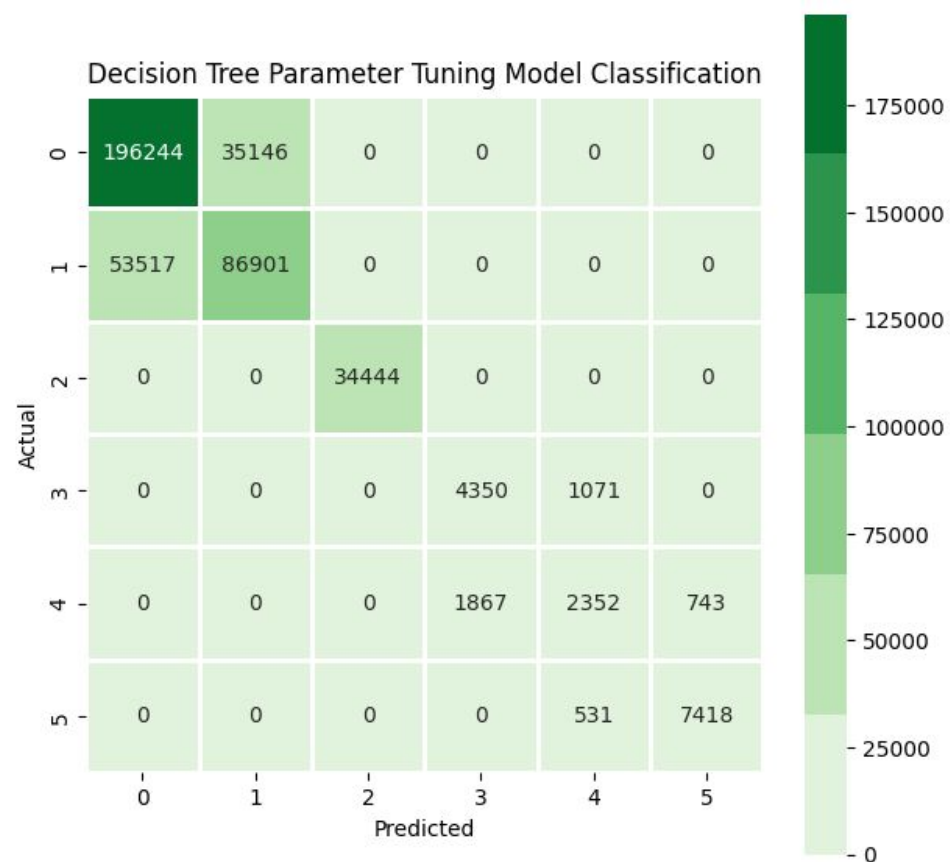
	weighted f1-score (regular ds)	weighted f1-score (undersampled ds)
Decision Tree	0.778	0.789
Random Forest	0.608	0.685
Boosted Trees	0.757	0.778

F1-SCORE COMPARISON BY DATASET

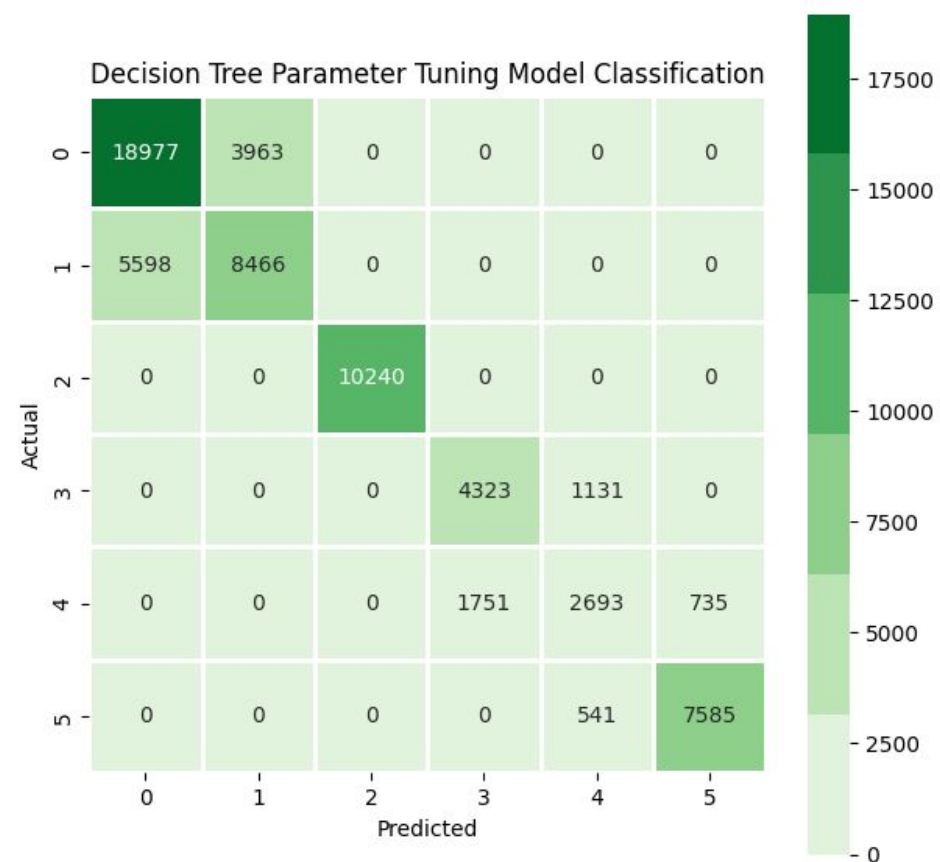


CLASSIFICATION COMPARISON BY CATEGORY

Original Dataset

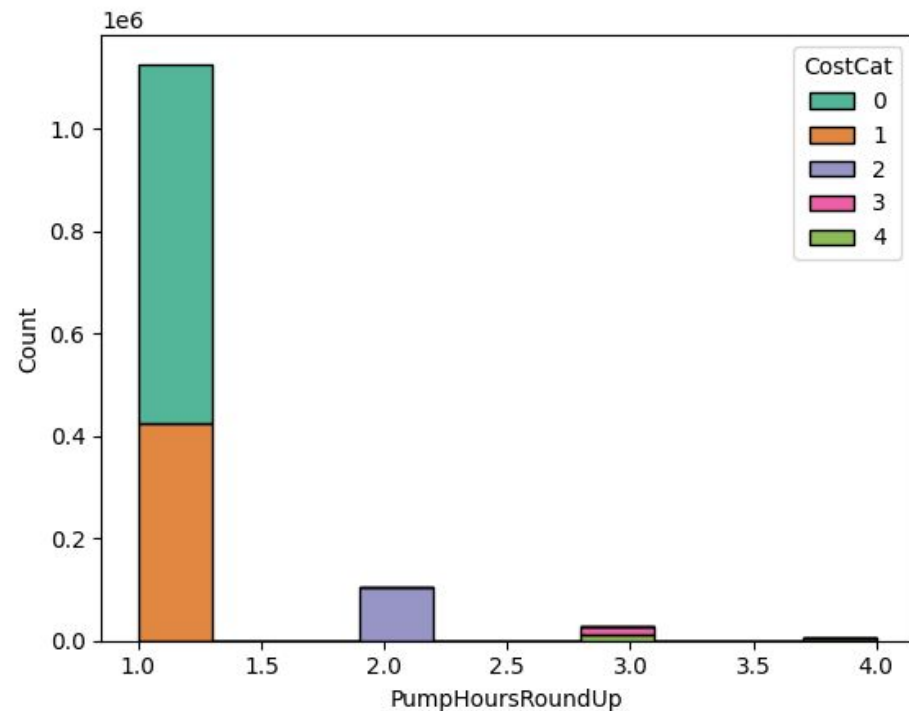


Undersampled Dataset

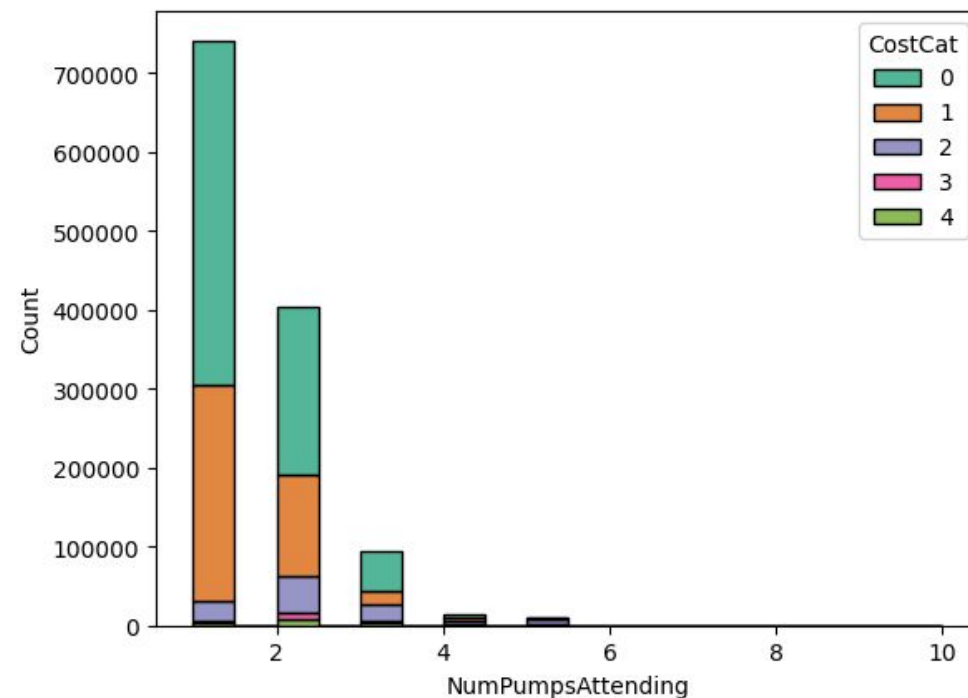


RELEVANT FEATURES FOR CLASSIFICATION

Pump Hours



Number of Pumps



CONCLUSION

1. Is it possible to **predict the approximate cost of firefighting operations** using tree-based models?
 - a. Yes, although not the best, the models produced acceptable performance
2. **What features correlate more or less** strongly with the operational cost?
 - a. Features relate to pump operations have high correlations while other variables like weather have weak correlation.
3. **Which tree-based model provides the best prediction** accuracy?
 - a. From our experiment, the best model is **decision tree**.

COMPARING WITH STATE OF THE ART

- Current industrial products: cost prediction for insurance and construction business
 - We found no solutions aimed at fire departments or public safety sector
- No accuracy data for comparison



ROOM FOR IMPROVEMENT

1. **Access to a variety of additional features**
 - a. Building materials
 - i. concrete vs. wooden walls
 - ii. old vs. renovated vs. new
 - b. Presence of fire sprinkler system
 - c. Neighborhood density
 - d. etc.
2. **Testing with different hyperparameters**
 - a. A broad range of hyperparameter values
 - b. Other unspecified parameters
3. **Test using additional datasets**
 - a. Other cities



THANK YOU 😊