L&T EduTech

Certificate in AI & Edge Computing for Industry Applications

PROJECT REPORT FOR

Problem Statement 2 (PS-2)
Garbage Bin Level Prediction and Collection

Khumanthem Catherine

VIT Bhopal University

Table of Contents

1. Student Information

Student Name: Khumanthem Catherine

University Name: VIT Bhopal University

University Registration Number: 23BEC10016

Submission Date: 7 July 2025

2. Introduction

2.1 Problem Statement

Urban areas are facing growing challenges in managing municipal solid waste effectively. The aim of this report is to explore the application of AI algorithms to improve waste collection by predicting when garbage bins will be full. Efficient management of waste is crucial for public health and the condition of cities. This involves collecting data from sensors or cameras on garbage bins, training machine learning models, and optimizing waste collection services.

2.2 Objective

The primary objectives are to:

- Predict the fill level of garbage bins using machine learning.
- Optimize waste collection schedules to minimize public health risks and improve sanitation.

3. Dataset Description

The datasets provided for this project include features such as:

- **Date and Time**: When the data was collected.
- **Fill Level**: Total litres and percentage of bin capacity.
- **Bin ID**: Identifier for the garbage bin.
- **Temperature**: Environmental conditions.
- **Battery Level**: To monitor sensor status.
- **Location**: Geographical data including latitude and longitude.

3.1 Data Attributes

- fill_levelin_litres: Continuous numeric field.
- fill_percentage: Percentage of bin fill capacity.
- temperature_in_ºc: Ambient temperature.
- battery_level: Level of battery in percentage.

4. Methodology

4.1 Data Preprocessing

- **Cleaning Column Names**: Standardizing naming conventions for easier access.
- **Data Type Conversion**: Ensuring that date, time, and various numeric fields are correctly formatted.
- **Handling Missing Values**: Filling NaNs with appropriate statistical methods.

4.2 Exploratory Data Analysis (EDA)

- **Distribution Plots**: Visualizing distributions of fill levels, percentages, and temperatures to understand their patterns.
- **Count Plots**: Observing the distribution of the target binary classification variable.

4.3 Feature Engineering

- **Creating Binary Indicators**: A binary target variable was created to indicate if the bin is full or not.

4.4 Model Development

Various machine learning models were evaluated, including:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- Gradient Boosting Classifier

- Support Vector Machine (SVM)

4.5 Model Evaluation Metrics

- **Precision**: Ratio of true positives to the sum of true positives and false positives.
- **Recall**: Ratio of true positives to the sum of true positives and false negatives.
- **F1-Score**: The harmonic mean of precision and recall.
- **Accuracy**: Proportion of true results in the total population.

4.6 Hyperparameter Tuning

Performed using GridSearchCV to optimize the Random Forest algorithm. The best-performing hyperparameters were identified.

5. Results

5.1 Model Performance

The models were evaluated using the test dataset. Key results are as follows:

- **Logistic Regression**:
  - Accuracy: 1.0000
  - Precision: 1.0000
  - Recall: 1.0000
  - F1-Score: 1.0000
- **Decision Tree**:
  - Accuracy: 1.0000
  - Precision: 1.0000
  - Recall: 1.0000
  - F1-Score: 1.0000
- **Random Forest**:
  - Best Parameters: {'max_depth': None, 'min_samples_split': 2, 'n_estimators': 50}
  - Accuracy: 1.0000

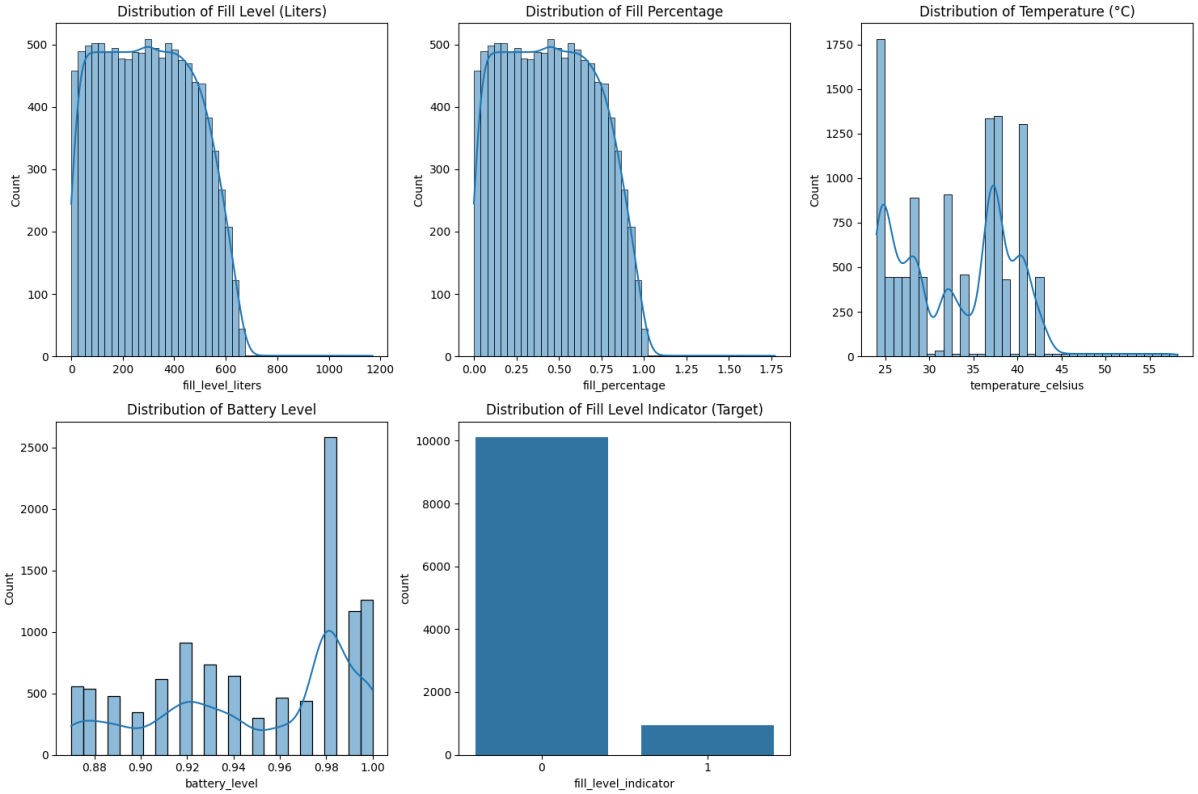
    - Precision: 1.0000
    - Recall: 1.0000
    - F1-Score: 1.0000
- **Gradient Boosting**:
    - Accuracy: 1.0000
    - Precision: 1.0000
    - Recall: 1.0000
    - F1-Score: 1.0000
- Support Vector Machine:
    - Accuracy: 0.9995
    - Precision: 1.0000
    - Recall: 0.9947
    - F1-Score: 0.9973

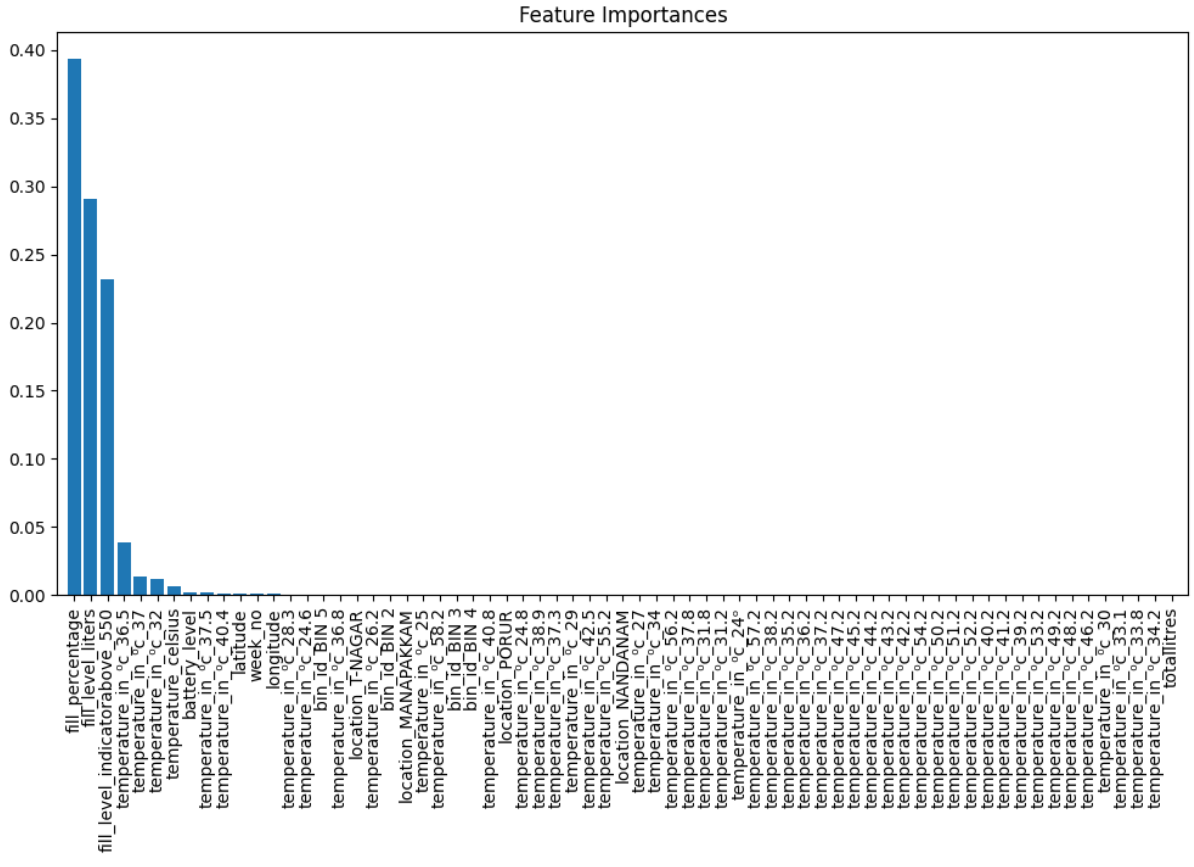
5.2 Feature Importance Analysis

The Random Forest model allowed for ranking feature importance:

1. fill_percentage: 39.33%
2. fill_level_liters: 29.10%
3. fill_level_indicatorabove_550: 23.14%

5.3 Visualization





6. Conclusion

The project successfully demonstrates the use of machine learning techniques for predicting garbage bin fill levels. The models achieved very high accuracy and can greatly enhance waste management strategies in urban settings by optimizing collection schedules.

7. Future Work

- Exploring additional models and techniques such as deep learning.
- Integrating real-time data for dynamic predictions.
- Expanding the datasets to include more data from different localities.
- Development of a mobile application to notify waste collection services based on predictive modelling results.