# Report

## Waste Management Using Machine Learning


**1. Abstract**

Waste management is a critical component of sustainable development. The process of classifying and sorting waste significantly impacts the efficiency of recycling and disposal. This project aims to leverage machine learning techniques to automate the classification of waste types using sensor data. A boosted random forest model was developed and fine-tuned, achieving high accuracy in classifying waste into predefined categories. The project also addressed key challenges such as class imbalance, data preprocessing, feature engineering, and dimensionality reduction, making the model suitable for deployment in real-world scenarios.


**2. Introduction**

**2.1 Background**

The rapid growth in urbanization and industrial activities has resulted in a significant increase in waste generation. Efficient waste management is essential to minimize environmental impact and promote recycling. However, manual sorting of waste is labor-intensive, time-consuming, and prone to human error.

**2.2 Problem Statement**

Existing waste classification systems are often limited by scalability and accuracy. With diverse types of waste materials and overlapping characteristics, an automated system is needed to ensure consistent and efficient classification.

**2.3 Objectives**

- Develop a machine learning model to classify waste based on sensor data.
- Improve classification accuracy by addressing class imbalance and enhancing feature engineering.
- Provide a scalable solution suitable for real-time deployment.

**2.4 Scope**

The scope of this project includes the analysis of sensor-based waste data, development of machine learning pipelines, and saving the final model for future integration with IoT systems.

**3. Methodology**

**3.1 Dataset**

- **Source**: A synthetic or real-world dataset comprising sensor readings for different waste types.
- **Features**:
  - **Inputs**: Inductive, capacitive, infrared, and moisture properties.
  - **Target**: Waste type (e.g., biodegradable, recyclable, hazardous).
- **Preprocessing**:
  - Removed irrelevant columns (e.g., timestamp).
  - Encoded categorical target variable using label encoding.

**3.2 Exploratory Data Analysis (EDA)**

- **Visualizations**:
  - **Bar Chart**: Shows the count of each waste type, highlighting data imbalance.
  - **Pie Chart**: Represents the proportional distribution of waste types.
- **Insights**:
  - Some waste types were significantly underrepresented, necessitating balancing techniques.

**3.3 Data Balancing**

- **Technique**:
  - Oversampling of minority classes using synthetic data generation or random resampling.
- **Outcome**:
  - Achieved a balanced dataset, ensuring the model does not favor majority classes.

**3.4 Feature Engineering**

- **New Features**:
  - inductive_to_capacitive: Ratio of inductive and capacitive properties.
  - infrared_to_moisture: Ratio of infrared and moisture properties.
- **Benefits**:
  - Enhanced feature space, improving the model's ability to distinguish waste types.

**3.5 Preprocessing**

- Applied standardization using StandardScaler to ensure uniform feature scaling.
- Reduced dimensionality using Principal Component Analysis (PCA) to improve computational efficiency and reduce noise.

## 4. Model Development

### 4.1 Model Selection

A boosted random forest classifier was chosen due to its ability to handle imbalanced datasets, its robustness to outliers, and its ensemble nature, combining multiple weak learners for better generalization.

### 4.2 Model Training

- **Training Process**:
  - Split the dataset into 80% training and 20% testing data.
  - Trained the model on the balanced dataset to ensure fairness.
- **Hyperparameters**:
  - Random Forest:
    - Number of estimators: 75
    - Maximum depth: 10
  - AdaBoost:
    - Number of estimators: 50
- **Algorithm**:
  - AdaBoost aggregates predictions from the random forest classifier to improve performance and reduce overfitting.

### 4.3 Model Evaluation

- **Metrics**:
  - Training and testing accuracy.
  - Precision, recall, and F1-score for each waste type to measure performance.
  - Confusion matrix to visualize correct and incorrect predictions.

\

## 5. Results and Discussion

### 5.1 Key Results

- The model achieved an accuracy of 65%
- High precision and recall for most waste types, indicating effective classification.

### 5.2 Observations

- Feature engineering played a significant role in improving model performance.
- Balancing the dataset ensured that the model was not biased toward majority classes.

- Dimensionality reduction (PCA) reduced computation time without significant loss of information.

### 5.3 Challenges

- **Sensor Noise**: The dataset might include noise from faulty sensors, impacting accuracy.
- **Real-Time Constraints**: Ensuring model deployment works seamlessly with IoT systems in real-time environments.

## 6. Deployment

- **Process**:
  - Saved the trained model using joblib for deployment in IoT systems.
  - The model can now classify waste types based on live sensor inputs.
- **Advantages**:
  - Scalable and efficient for integration in waste management facilities.
- **Future Improvements**:
  - Expand the dataset with real-world sensor data to enhance robustness.
  - Implement advanced algorithms like convolutional neural networks (CNNs) for image-based waste classification.

## 7. Conclusion

This project demonstrated the effectiveness of machine learning in automating waste classification. By addressing class imbalance and leveraging feature engineering, the model achieved high accuracy and fairness. With potential for real-world deployment, this solution can significantly improve waste sorting processes, enhancing sustainability efforts.