Waste Management and Recycling in Indian Cities

(Machine Learning-based Urban Sustainability Project)

Submitted by:
Raj Chhapariya
B. Tech in Information Technology

Date: October 15, 2025

GitHub Repository:
https://github.com/Rajchhapariya/Waste-Management-and-Recycling-in-Indian-Cities

Deployed Application:
https://hackathon-deploy-indian-waste.onrender.com/

1. Introduction

India’s rapid urbanization has led to a massive increase in municipal solid waste generation. According to recent estimates, **urban India produces over 150,000 tonnes of solid waste per day**, with only a portion being scientifically processed or recycled.
Inefficient segregation, collection, and disposal methods have made **waste management a critical challenge** for sustainable urban development.

This project aims to leverage **data analytics and machine learning** to analyze waste generation patterns, predict recycling potential, and assist local bodies in improving collection and recycling efficiency across Indian cities.

---

2. Methodology

2.1 Data Preprocessing

- **Data Sources:**
    - City-wise waste collection data (e.g., Swachh Bharat Mission datasets, CPCB reports).
    - Recycling statistics from municipal corporations.
    - Demographic and geographic data (e.g., population density, area, region type).
- **Data Cleaning:**
    - Separated two columns Latitude and Longitude from a single column.
    - Dropped a column landfill name which was of no use in model training.
- **Encoding:**
    - One-hot encoding for various columns such as City, Waste type and Disposal method.

2.2 Feature Engineering

- **Numerical Features:**
    - Chose numerical features based on their correlation with the target variable that was Recycling Rate (%).

- **Categorical Features:**
    - Converted categorical features such as City, Waste type and Disposal method to numerical feature as it was necessary for model training because Machine needs numerical data to train the model.

2.3 Model Selection

- For **regression** (predicting recycling rate):
    - Linear Regression, Random Forest Regressor, XGBoost Regressor, LGBM Regressor.
- **Final Model:** Linear Regression Model chosen because of lowest RMSE value among all the four models.

2.4 Deployment Procedure

- Created a **Flask web interface** where users can input a csv file which has city-level data.
- The backend model predicts recycling rate percentage.
- Deployed the flask app on Render.

---

3. Results

3.1 Model Performance

Metric	Value
RMSE	17.5686

3.2 Visualizations

- **City-wise Waste Generation:** Bar chart comparing total waste per city.
- **Trend Analysis:** Line plot showing seasonal waste variation across cities.
- **Feature Importance:** Bar chart displaying the most influential predictors.

3.3 Insights

- Cities like Jaipur, Ranchi, Lucknow, Mumbai generate the highest waste volumes.
- The waste generation by all the cities are almost uniform.
- All the waste types equally contributes significantly to landfill accumulation.

---

4. Discussion

4.1 Challenges

- Lack of uniform data collection standards across municipal bodies.
- Missing geolocation or temporal data for smaller cities.
- Variations in classification schemes for waste types.

4.2 Limitations & Improvements

- Dataset limited to select Indian cities — expanding to more cities would improve generalization.
- Real-time IoT integration (smart bins, GPS trucks) can provide live data streams.
- Future models can incorporate **deep learning** for better pattern recognition.

4.3 Real-world Implications

- Enables data-driven waste collection scheduling for municipal corporations.
- Supports government schemes like **Swachh Bharat Mission 2.0**.
- Reduces landfill load by identifying recycling hotspots.
- Promotes citizen awareness through predictive insights and visual dashboards.

4.4 Future Scope

- Development of a **real-time waste monitoring system** using sensors.
- Policy-level integration to guide recycling infrastructure investment.
- Expansion into **circular economy analytics** for reuse and composting opportunities.

---

5. Conclusion

This project demonstrates that applying **machine learning to urban waste data** can uncover valuable insights for improving recycling rates and optimizing waste management systems in Indian cities.
By identifying trends, inefficiencies, and opportunities for intervention, this approach contributes to **India's vision of sustainable urban development** and **zero landfill cities**.