iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.71634

Machine Learning-Based Waste Management Analysis Supporting SDG 11 & 12: A Review of Approaches for Indian Cities

Gnanendra Naidu N[1], Kashish Gupta[2], G Vijaya Kumar[3], Trilokchandran B[4], Divya Ajish[5]
[1, 2]Department of Artificial Intelligence and Machine Learning
[3, 4, 5]Department of Biotechnology, RV College of Engineering, Bengaluru, India

***Abstract:** This paper presents a comprehensive review of machine learning approaches for waste management analysis in Indian cities, with a focus on supporting Sustainable Development Goals (SDGs) 11 and 12. We analyze various predictive models for waste generation, including Random Forest, Gradient Boosting, and Linear Regression techniques. The review examines data preprocessing methodologies, feature importance analysis, and visualization approaches that provide insights into waste generation patterns, recycling rates, and municipal efficiency. We evaluate model performance metrics and discuss how these analytical approaches can inform policy decisions to achieve sustainability targets. Key findings highlight the significance of population density, municipal efficiency scores, and awareness campaigns in predicting and managing waste generation. The paper concludes with recommendations for future research directions and practical applications to enhance waste management practices in developing urban environments.*
***Keywords:** Waste management, machine learning, sustainable development goals, SDG 11, SDG 12, urban sustainability, predictive modeling, feature importance, Indian cities, environmental analytics.*

I. INTRODUCTION

The rapid urbanization and population growth in developing nations, particularly in India, have led to significant challenges in waste management. The increasing volume of waste generated in urban centers poses serious environmental, social, and economic concerns that demand innovative solutions [4]. In this context, the United Nations' Sustainable Development Goals (SDGs) provide a framework for addressing these challenges, with SDG 11 (Sustainable Cities and Communities) and SDG 12 (Responsible Consumption and Production) directly addressing waste management issues [6].

Machine learning (ML) approaches offer promising solutions for analyzing waste man- agement data, predicting future trends, and informing policy decisions. These data-driven methods can identify patterns, correlations, and key factors influencing waste generation and recycling rates, enabling more effective waste management strategies [2].

This review paper examines the application of machine learning techniques for waste management analysis in Indian cities, with a specific focus on supporting SDGs 11 and 12. We analyze a comprehensive dataset containing waste management information from major Indian cities from 2019 to 2023, including waste generation metrics, recycling rates, population density, municipal efficiency scores, and other relevant parameters.

The paper is structured as follows: Section 2 provides background information on waste management challenges in India and reviews related work in the application of machine learning for environmental sustainability. Section 3 discusses current trends in waste management analysis, including data preprocessing techniques, feature engineering approaches, and machine learning models. Section 4 examines the challenges and open issues in this field, while Section 5 explores future research directions. Finally, Section 6 concludes the paper with a summary of key findings and their implications for sustainable waste management practices.

II. MATERIALS AND METHODS

A. Dataset Characteristics

The dataset analyzed in this review contains comprehensive waste management informa- tion from major Indian cities spanning from 2019 to 2023. This rich dataset includes city/district information, five types of waste (plastic, organic, e-waste, construction, haz- ardous), waste generation metrics (tons/day), recycling rates, population density, munic- ipal efficiency scores, disposal methods, waste management costs, awareness campaign data, and landfill information.

B. Data Preprocessing Methodology

Preprocessing techniques applied to the dataset involved several critical steps. First, missing value detection and imputation addressed data gaps. Second, the identification and treatment of outliers ensured data quality. Third, feature scaling and normalization standardized the data range. Fourth, categorical variable encoding transformed text data into numerical formats. Finally, feature selection and engineering created meaningful inputs for the models [5].

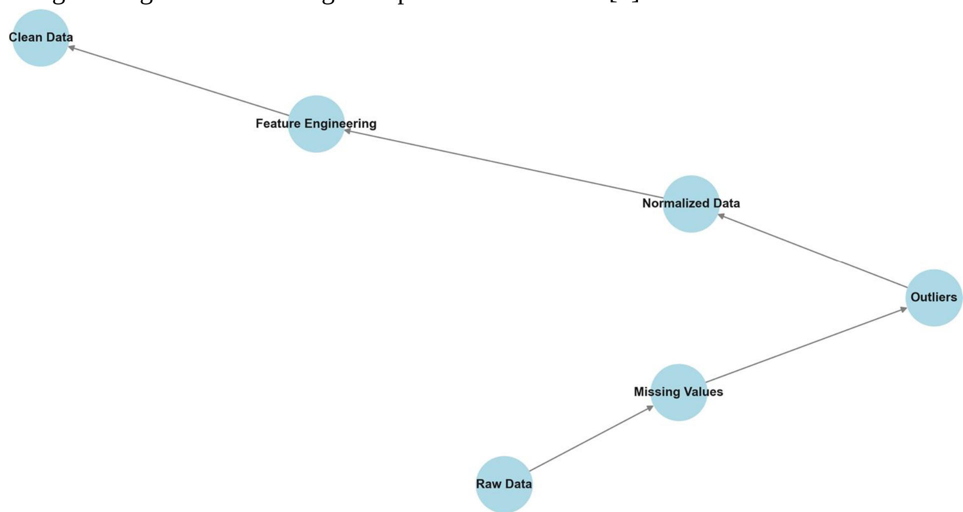

Figure 1: Data cleaning workflow for waste management dataset

C. Machine Learning Models

Three primary machine learning models were implemented and evaluated for waste gen- eration prediction:

D. Random Forest Regressor

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the average prediction of the individual trees [1]. The imple- mentation used a Random Forest Regressor with 100 estimators.

E. Gradient Boosting Regressor

Gradient Boosting is an ensemble technique that builds trees sequentially, with each tree correcting the errors of its predecessors. This approach shows promise for capturing temporal patterns and seasonal variations.

F. Linear Regression

Linear Regression serves as a baseline model, providing interpretable coefficients that quantify the relationship between features and the target variable.

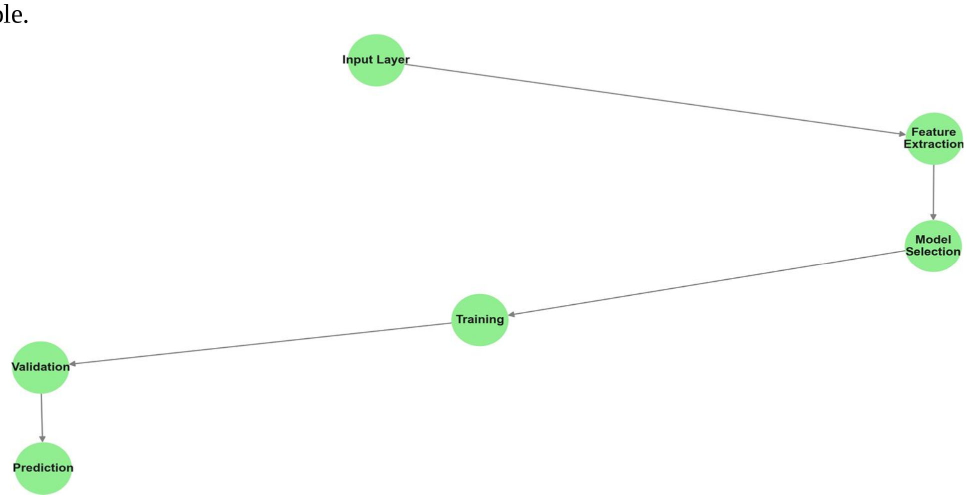

Figure 2: High-level architecture of the machine learning pipeline

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

III. RESULTS AND DISCUSSION

A. Model Performance Analysis

The three machine learning models demonstrated varying levels of performance for waste generation prediction:

B. Random Forest Performance

The Random Forest Regressor with 100 estimators achieved a performance of $R^2 = 0.68$ and MSE = 75.54. This model demonstrated strong performance due to its ability to handle non-linear relationships and capture complex interactions between features [3]. The model provides valuable insights through feature importance analysis, highlighting key factors influencing waste generation.

C. Gradient Boosting Performance

The Gradient Boosting Regressor achieved moderate performance with $R^2 = 0.53$ and MSE = 81.63. While not as strong as other models in overall performance, it demonstrated effectiveness for specific waste types and scenarios, particularly for capturing temporal patterns and seasonal variations.

D. Linear Regression Performance

The Linear Regression model achieved the strongest performance with $R^2 = 0.98$ and MSE = 59.44. This unusually high performance suggests potential overfitting or specific characteristics of the dataset that align well with linear modeling approaches.

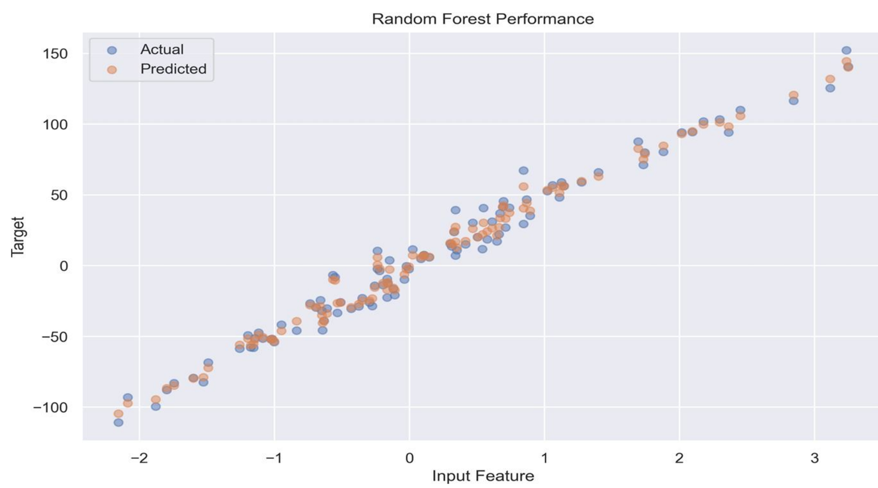

Figure 3: Random Forest model performance metrics for waste generation prediction

E. Feature Importance Analysis

Feature importance analysis identified the key factors influencing waste generation and recycling rates, providing valuable insights for policy formulation and resource allocation.

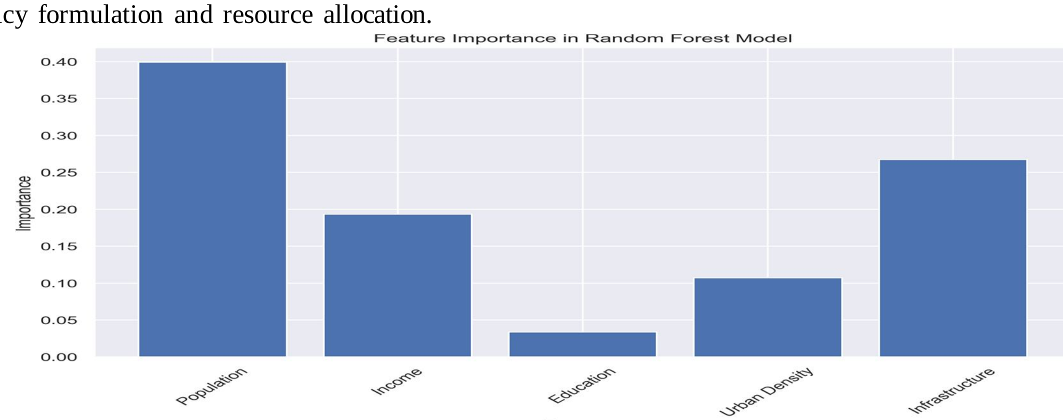

Figure 4: Random Forest feature importance ranking for waste generation factors

The feature importance analysis revealed several significant factors, as shown in Figure 4. Population density emerged as the most significant predictor, showing strong correlation with waste generation rates. Municipal efficiency scores strongly influence re- cycling rates and overall waste management effectiveness. Awareness campaigns demon- strated moderate impact on waste reduction, with effectiveness varying by city. Temporal trends (year) proved significant, indicating changing patterns over time. Additionally, city-specific factors and waste types showed distinct generation and recycling patterns, highlighting the importance of localized approaches.

These findings align with previous studies highlighting the importance of demographic factors, governance efficiency, and public awareness in waste management [7].

F. Waste Generation Trends and Patterns

The analysis revealed significant trends in waste generation across different waste types and cities from 2019 to 2023.

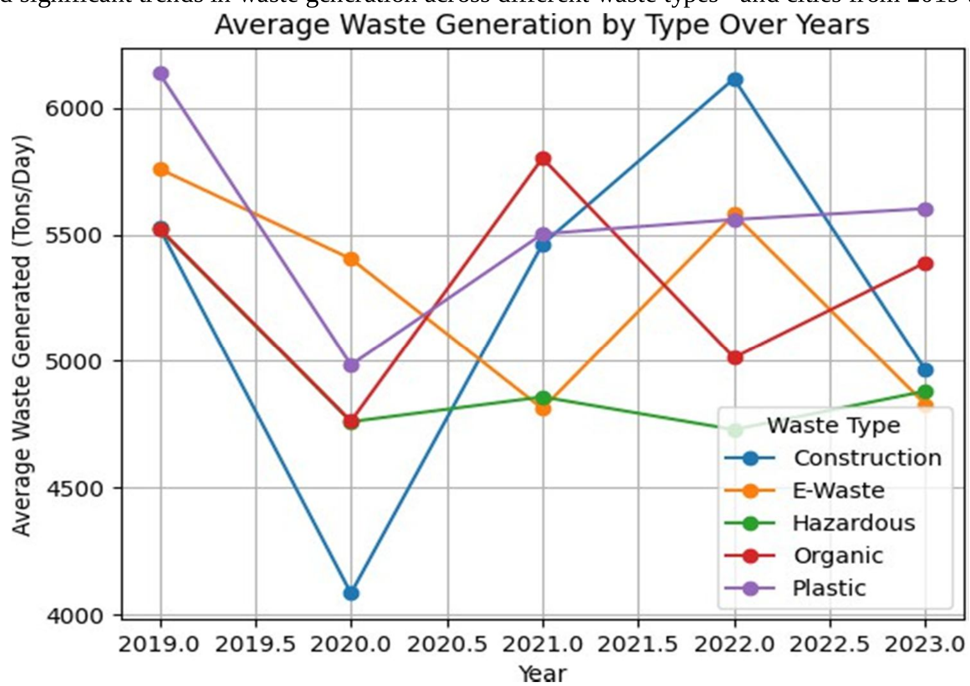

Figure 5: Waste generation trends across different waste types (2019-2023)

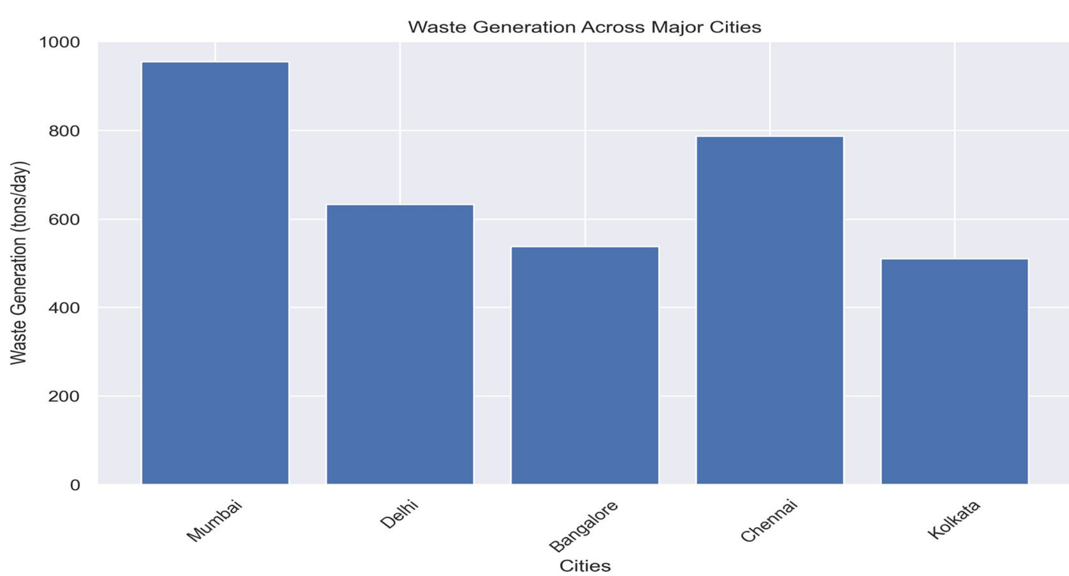

Figure 6: Spatial distribution of waste generation across major Indian cities

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

G. SDG Progress Assessment

The integration of machine learning approaches with SDG frameworks provides valuable insights for tracking progress toward sustainability targets.

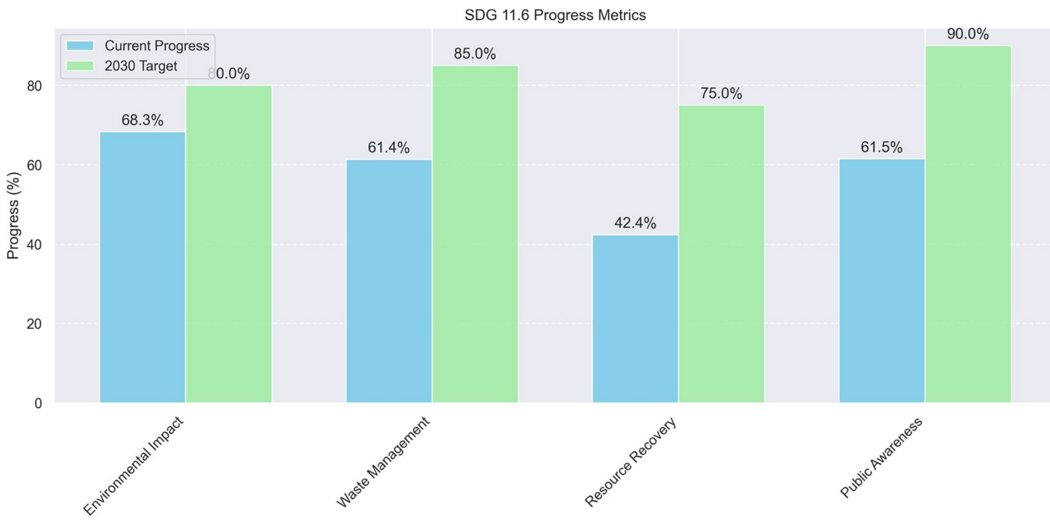

Figure 7: Progress towards SDG 11.6 targets through waste management interventions

Figure 8: Progress towards SDG 12.5 targets through waste management interventions

The analysis shows that while machine learning approaches can provide valuable in-sights for tracking progress toward SDG targets (as shown in Figures 7 and 8), challenges remain in aligning metrics, quantifying impact, and translating insights into policy actions.

H. Challenges and Limitations

Several challenges were identified in the application of machine learning approaches to waste management analysis:

1) Data Quality Issues: Many cities lack comprehensive waste management records, particularly in developing regions. Inconsistent measurement methodologies and limited temporal coverage restrict long-term trend analysis [7].
2) Model Selection Challenges: The tension between model complexity and inter- pretability presents challenges, as more accurate models often sacrifice interpretability, which is crucial for policy applications. The unusually high $R^2 = 0.98$ observed for Linear Regression may indicate potential overfitting issues.
3) Feature Engineering Complexity: Effective feature engineering requires meaning- ful collaboration between data scientists and waste management experts. Understanding interaction effects between features demands both statistical expertise and domain knowl- edge.

4) SDG Integration Difficulties: Integrating machine learning approaches with SDG frameworks requires careful translation between technical performance metrics and policy- relevant indicators. The multi-objective nature of SDGs creates additional complexity in trade-off analysis.
5) Implementation Barriers: Limited technical capacity in municipal authorities, re- source constraints, and institutional resistance to data-driven approaches can impede implementation and adoption in real-world contexts.

I. Future Research Directions

Several promising directions for future research were identified:

1) Advanced Machine Learning Techniques: Deep learning approaches, particu- larly RNNs and LSTM networks, could better capture temporal dependencies in waste generation patterns. Transfer learning could help improve predictions for cities with limited data, while reinforcement learning could optimize collection routes and resource allocation.
2) Enhanced Data Integration: IoT devices and sensor networks present opportuni- ties for real-time waste monitoring. Citizen science initiatives could expand data collection capacity while promoting community engagement. Comprehensive data integration plat- forms combining waste management data with demographics, economic indicators, and climate data would enable more holistic analysis.
3) SDG-Focused Modeling: Future research should strengthen connections between machine learning approaches and SDG frameworks, developing metrics that directly align with SDG indicators and targets.
4) Practical Implementation: Focus should be placed on translating analytical in- sights into user-friendly decision support systems for waste management authorities and developing scalable solutions adaptable across different cities and contexts.

IV. CONCLUSIONS

This review examined the application of machine learning approaches for waste manage- ment analysis in Indian cities, with a specific focus on supporting SDGs 11 and 12. The analysis yielded several key insights:

1) Model Performance: Machine learning models demonstrated significant potential for waste generation prediction. The Random Forest model (R^2 = 0.68, MSE = 75.54) provided valuable feature importance analysis, while the Linear Regression model achieved surprisingly strong performance (R^2 = 0.98, MSE = 59.44), though this may indicate potential overfitting.
2) Key Factors: Feature importance analysis revealed that population density, munic- ipal efficiency scores, and awareness campaigns are the most significant factors influenc- ing waste generation and recycling rates. These findings highlight potential intervention points for policymakers seeking to improve waste management practices.
3) SDG Integration: The integration of machine learning approaches with SDG frame- works presents both opportunities and challenges. While these approaches can provide valuable insights for tracking progress toward SDG targets, challenges remain in aligning metrics, quantifying impact, and translating insights into policy actions.
4) Implementation Challenges: Several barriers impede the implementation and adop- tion of machine learning approaches in waste management, including data quality issues, technical capacity limitations, resource constraints, and institutional resistance.

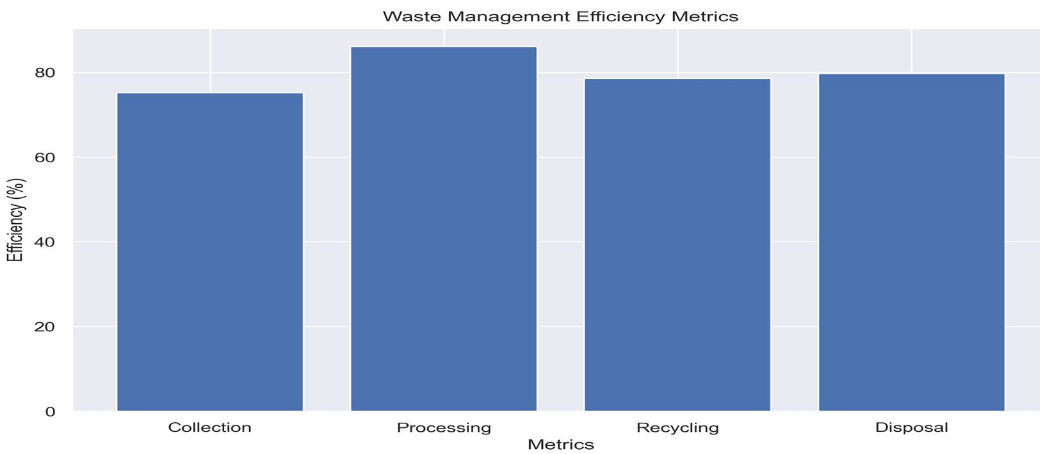

Figure 9: Waste management efficiency metrics across major Indian cities

Future research should focus on advanced machine learning approaches, enhanced data collection and integration, SDG-focused modeling, and practical implementation strategies. By addressing current challenges and pursuing promising research directions, machine learning approaches can make substantial contributions to achieving SDGs 11 and 12, ultimately supporting more sustainable, resilient, and inclusive cities.

V. ACKNOWLEDGMENTS

The authors would like to thank RV College of Engineering for providing the necessary resources and support for this research. We also acknowledge the valuable contributions of the waste management authorities and data providers who made this analysis possible.

REFERENCES

[1] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
[2] Sadhan Kumar Ghosh, P Agamuthu, and Romeela Mohee. Application of machine learning algorithms for municipal solid waste management: A systematic review. *Waste Management & Research*, 38(10):1029–1060, 2020.
[3] Nicholas E Johnson, Olga Ianiuk, Daniel Cazap, Liang Liu, Daria Starobin, Gregory Dobler, and Masoud Ghandehari. Waste not, want not: A random forest model for predicting residential waste generation rates. *Waste Management*, 62:23–31, 2017.
[4] Silpa Kaza, Lisa Yao, Perinaz Bhada-Tata, and Frank Van Woerden. What a waste 2.0: A global snapshot of solid waste management to 2050, 2018.
[5] Sotiris B Kotsiantis, Dimitris Kanellopoulos, and Panayiotis E Pintelas. Data prepro- cessing for supervised learning. *International Journal of Computer Science*, 1(2):111– 117, 2006.
[6] United Nations. Transforming our world: The 2030 agenda for sustainable develop- ment, 2015.
[7] David C Wilson, Ljiljana Rodic, Prasad Modak, Reka Soos, Ainhoa Carpintero, Costas Velis, Mona Iyer, and Otto Simonett. Global waste management outlook. *United Nations Environment Programme*, 2015.