International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, July-Aug-2026, ISSN (Online): 2395-566X

A Machine Learning Framework for Predicting E-Waste Generation Hotspots in Tier-2/3 Indian Cities: An Extended Formulation and Simulation-Based Validation Study

Prof. Priyanka Mahajan
Basic Engineering Science
Indira College of Engineering and Management, Pune, India

Rohit Balaji Pingale
Department of Information Technology
Indira College of Engineering and Management, Pune, India

Vaishnavi Laxman Shinde
Department of Information Technology
Indira College of Engineering and Management, Pune, India

Abstract — A India's electronic waste (e-waste) generation has increased rapidly, while formal collection and recycling infrastructure remains concentrated in major metropolitan areas. Tier-2 and Tier-3 cities continue to face challenges in planning efficient collection networks due to limited data-driven decision support. This paper presents a machine learning-based framework for estimating ward-level e-waste generation using publicly available demographic, economic, and administrative indicators. A Random Forest regressor is proposed as the primary prediction model and is benchmarked against Gradient Boosting and Ordinary Least Squares regression to evaluate predictive performance. The framework incorporates a reproducible data-engineering pipeline integrating data from the Central Pollution Control Board (CPCB), State Pollution Control Boards (SPCBs), Census records, and Urban Local Bodies (ULBs). To validate the proposed workflow before large-scale field deployment, a clearly labelled synthetic dataset is developed for model training and evaluation. Feature importance analysis is further employed to identify the key factors influencing e-waste generation, enabling practical insights for municipal authorities and Extended Producer Responsibility (EPR) stakeholders. The proposed approach offers a scalable, low-cost, and replicable methodology for supporting sustainable e-waste management, optimizing collection planning, and reducing the infrastructure gap between metropolitan and emerging urban regions in India.

Keywords— e-waste management; machine learning; Extended Producer Responsibility (EPR); Tier-2/3 cities; Random Forest regression; predictive analytics; municipal solid waste; sustainable urban planning

I. INTRODUCTION

Electronic waste is, by most national accounts, the fastest-growing waste stream in India. Figures published by the Central Pollution Control Board (CPCB) place aggregate generation across twenty-one notified categories of electrical and electronic equipment at 13,46,496.31 tonnes in FY 2020-21, rising to 16,01,155.36 tonnes the following year. The regulatory response has kept pace on paper: the Ministry of Environment, Forest and Climate Change replaced the 2016 rules with the E-Waste (Management) Rules, 2022, effective April 2023, which introduced a tradable Extended Producer Responsibility (EPR) certificate regime. As of December 2024 the formal ecosystem under these rules comprised more than 7,200 registered producers, roughly 295 recyclers, 35 manufacturers and 53 refurbishers..

What the national numbers obscure is where the waste is actually generated and where formal capacity to handle it exists. Both quantities are unevenly distributed across the urban hierarchy. Early ecosystem studies of Bengaluru already documented a persistent gap between generation and formal processing capacity in a metro with comparatively mature infrastructure, and subsequent national reviews of the regulatory landscape describe awareness, collection density and recycler presence as skewed toward a small number of large cities. Parallel work on the technology side has focused on automating the identification of e-waste once it has entered

a collection stream — for example, robotic sorting attached to municipal garbage trucks — but such systems have so far only been piloted in large urban settings and, more fundamentally, presuppose that collection infrastructure already exists at the point of interest.

Tier-2 and Tier-3 Indian cities sit on the wrong side of both gaps. Electronics penetration in these cities is rising with expanding middle-class consumption, but end-of-life appliances are still routed predominantly through informal aggregators and scrap dealers rather than EPR-registered recyclers. Municipal bodies in these cities are consequently asked to plan collection-point networks and support producer compliance without the demand-forecasting tools that metro administrations increasingly rely on, while recyclers looking to expand beyond saturated metro markets have limited visibility into where non-metro collection investment would actually pay off.

Figure 1. An informal e-waste collection/dismantling site in a Tier-2/3 Indian city.

This paper addresses that planning gap directly. We propose, and in this extended version fully specify, a machine-learning framework that estimates ward- and city-level e-waste generation from indicators that are already published by government sources — population density, income proxies, historical consumer-durable ownership, urbanisation rate and literacy rate — so that the resulting tool remains reproducible by ULBs and recyclers who do not have dedicated data-science teams. Relative to the initial proposal, the present paper makes four additional contributions:

- a formal mathematical statement of the estimation problem and of the three candidate estimators (Random Forest, Gradient Boosting, ordinary least squares), including the metrics used to compare them;
- a data-engineering pipeline that maps each candidate feature to a specific, named government source and describes the joins and cleaning steps needed to assemble a ward-level training table;
- a synthetic simulation study, explicitly labelled as such, that exercises the entire pipeline end-to-end so that the evaluation code, cross-validation protocol and reporting format are validated before real ward-level data are in hand; and
- a discussion that translates model outputs into specific planning actions for ULBs and recyclers operating under the EPR regime.

The remainder of the paper is organised as follows. Section II reviews related work in machine-learning-based waste forecasting and in India-specific e-waste policy research. Section III formalises the prediction problem. Section IV describes the proposed framework in full — data sources, feature engineering, model formulation and validation strategy. Section V describes the study area and the synthetic experimental setup used to validate the pipeline in the absence of field data. Section VI reports illustrative results and their interpretation. Section VII discusses policy implications, Section VIII discusses limitations and threats to validity, and Section IX concludes with a concrete field-validation plan

II. LITERATURE REVIEW

1. Machine Learning in Municipal Solid Waste Management

Applications of machine learning and deep learning across the municipal solid-waste pipeline now span waste-characteristics forecasting, bin-fill-level monitoring and collection-route optimisation. Comparative surveys of this literature report performance that is promising in aggregate but highly sensitive to data quality and to how the underlying prediction problem is framed, with accuracy varying considerably across cities and waste categories depending on feature availability. This sensitivity is a central design constraint for the present work: any framework intended for Tier-2/3 ULBs must be robust to noisy, incomplete or infrequently updated administrative data, since these cities do not have the sensor networks or high-frequency reporting available in metro pilot studies.

2. AI for E-Waste Identification, Sorting and Collection

Within the narrower e-waste sub-domain, published work concentrates heavily on item-level identification and sorting rather than on forecasting how much e-waste a given locality will generate. Madhav et al. proposed a mobile robot mounted on municipal garbage trucks that applies transfer learning to identify and segregate household e-waste at the point of collection, reporting classification accuracy above 96% with a modified ResNet-50 backbone. This class of system is valuable

for improving material recovery once e-waste has entered the collection stream, but it takes the existence and location of that stream as given; it offers no guidance on where a collection stream should be established in the first place, which is precisely the decision Tier-2/3 ULBs currently lack support for.

3. India-Specific E-Waste Policy Research

India-specific policy scholarship supplies useful context without closing this gap. Borthakur and Govind's ecosystem study of Bengaluru documents a substantial and persistent divide between the city's informal and formal recycling sectors despite comparatively developed infrastructure. Arya and Kumar provide a national synthesis of India's e-waste regulatory evolution and highlight recurring weaknesses in awareness, enforcement and the geographic distribution of infrastructure across city tiers. The Global E-waste Monitor similarly situates India's growth in generation within a broader global pattern in which formal collection rates lag generation growth in most middle-income economies, reinforcing that this is not solely an Indian policy failure but a structural feature of rapid electronics adoption outpacing recycling infrastructure build-out. None of these studies, however, proposes a predictive, data-driven tool that a non-metro municipal body could apply directly to its own ward-level data.

4. Research Gap

Two gaps persist once this literature is read together. First, the great majority of AI-based e-waste research targets classification and sorting of waste that has already entered the collection stream, rather than forecasting generation volumes ahead of infrastructure investment. Second, the limited India-specific empirical work on infrastructure gaps has examined individual metro case studies — Bengaluru in particular — rather than proposing a replicable, low-cost predictive methodology that resource-constrained Tier-2/3 ULBs could realistically adopt. The framework developed in this paper is designed specifically to close both gaps simultaneously.

III. PROBLEM FORMULATION

We treat ward-level e-waste generation as a supervised regression problem. Let a city be partitioned into wards indexed $i = 1, \ldots, n$. For each ward i we observe a feature vector $x_i = (x_{i1}, \ldots, x_{ip})$ drawn from publicly available demographic, economic and administrative sources (Section IV-B), and we wish to estimate the annual e-waste generation volume y_i (in kilograms, normalised by ward population where appropriate for cross-city comparison).

The learning task is to recover a function f such that $\hat{y}_i = f(x_i)$ minimises expected prediction error over wards not seen during training. Because ground-truth ward-level generation volumes are not directly metered anywhere in India — CPCB and SPCB statistics are reported at city, state or producer level, not at ward level — y_i is itself a latent quantity that must eventually be approximated through the ground-truthing exercises described in Section

VIII. The framework in this paper is therefore best understood as a mapping from observable, low-cost covariates to a hotspot ranking of wards, rather than as a claim to recover an absolute tonnage figure with metrological precision. Wards are ranked by $\hat{y}_i$ to produce a prioritised list for collection-point placement, and the same ranking is aggregated to city level to compare Nashik, Kolhapur, Chhatrapati Sambhajinagar (Aurangabad) and Nagpur against one another.

Three candidate forms for f are evaluated in this paper: an ordinary least-squares linear model as a transparent baseline, a Random Forest regressor as the primary proposed estimator, and a Gradient Boosting regressor as a stronger but less interpretable benchmark. All three are trained on the same feature matrix $X \in \mathbb{R}^{n \times p}$ and compared using the validation protocol in Section IV-E.

IV. PROPOSED FRAMEWORK

1. System Architecture Overview

The framework is organised as a five-stage pipeline, summarised in Table I. Government data sources are ingested and joined at ward level; a feature-engineering stage derives the five candidate predictors described below; the resulting feature matrix is passed to the three candidate regressors; the best-performing model produces a city-wise, ward-ranked e-waste hotspot map; and that map feeds directly into municipal EPR collection-point planning and into recyclers' partnership decisions.

Fig. 1. Five-stage pipeline realising the proposed framework

2. Data Sources and Preprocessing

The framework draws exclusively on publicly available secondary data so that the methodology remains low-cost and replicable by ULBs without dedicated analytics staff. Three source families are used. CPCB e-waste inventory statistics and annual reports supply national- and state-level generation totals used to calibrate city-level scaling factors. State Pollution Control Board (SPCB) registration and processing records

supply the count and location of registered dismantlers, recyclers and collection points already operating in each study-area city, which serve both as a feature (existing formal capacity) and as a partial check on plausibility of predicted hotspots. Census of India and Urban Local Body records supply ward-level demographic and socio-economic indicators — population, household counts, literacy rate and consumer-durable ownership from the most recent housing and amenities tables — that form the backbone of the feature set. Because these sources are collected on different cycles and at different administrative resolutions, a preprocessing stage aligns them to a common ward boundary definition, forward-fills or interpolates indicators that are only available at less frequent census intervals, and flags wards with more than a threshold proportion of missing fields for exclusion or imputation rather than silent inclusion.

3. Feature Engineering

Five candidate predictive features are carried forward from the original proposal and given precise operational definitions here so that another team could reconstruct the same feature table from public sources without ambiguity:

- Population density — ward population divided by ward area (persons/km²), from Census/ULB records.
- Per-capita income proxy — district-level Gross Domestic Product (GDP) divided by district population, apportioned to wards; used because household income is not published at ward level.
- Historical consumer-durable ownership trend — the year-on-year change in the share of households reporting ownership of televisions, computers, mobile phones and refrigerators in successive housing and amenities surveys, used as a proxy for electronics penetration and hence future waste generation.
- Urbanisation rate — the ward's share of built-up area or its classification trajectory from rural to urban/municipal status over the preceding decade.
- Literacy rate — used as a proxy for awareness of formal e-waste disposal channels, following the association between education level and participation in formal recycling programmes reported in the broader waste-management literature.

Feature-importance rankings produced by the trained Random Forest and Gradient Boosting models (Section VI) are used post hoc to prune or re-weight this list; the five features above should therefore be read as the candidate set entering training rather than as a final, fixed specification.

4. Model Formulation

Let $X \in \mathbb{R}^{n \times 5}$ be the ward-by-feature matrix and $y \in \mathbb{R}^n$ the (latent, to be ground-truthed) generation target. The linear baseline estimates $\hat{y} = X\beta + \varepsilon$ by ordinary least squares, minimising the residual sum of squares $\Sigma_i(y_i - x_i^T\beta)^2$. This baseline is retained deliberately: its coefficients are directly interpretable by non-specialist ULB staff and it quantifies how much predictive value the ensemble methods actually add over a transparent linear model.

The Random Forest regressor, the paper's primary proposed estimator, builds an ensemble of B regression trees {T₁, …, T_B}, each trained on a bootstrap resample of the ward set with a random subset of features considered at each split, and predicts $\hat{y}_i = (1/B)\ \Sigma\_b\ T\_b(x_i)$. Random Forests were selected over a single decision tree for two reasons specific to this problem: bootstrap aggregation reduces the variance that would otherwise result from the small, noisy, and unevenly updated ward-level tables typical of non-metro administrative data, and the model's permutation- or impurity-based feature importances give ULB planners an interpretable ranking of which indicators are driving a ward's predicted hotspot status.
The Gradient Boosting regressor is included as a stronger benchmark. It fits an additive sequence of shallow trees, $F\_m(x) = F\_\{m-1\}(x) + \eta \cdot h\_m(x)$, where each h_m is trained on the negative gradient (pseudo-residuals) of the loss with respect to $F\_\{m-1\}$ and η is a shrinkage/learning-rate parameter, following the general formulation of gradient boosting machines. Comparing Random Forest against Gradient Boosting quantifies the marginal benefit of sequential error-correction over independent bagged trees on this specific, feature-sparse problem, rather than assuming ensemble sophistication automatically improves on a simpler bagged model — a distinction that matters because Gradient Boosting's added variance-reduction typically requires more, not less, data than Tier-2/3 ULBs are likely to have available.

5. Validation Strategy

All three models are compared using k-fold cross-validation with k = 5, so that each ward appears in exactly one held-out fold across the five training/testing splits. Three metrics are reported for each model: Root Mean Squared Error, $RMSE = \sqrt{[(1/n)\ \Sigma_i\ (y_i - \hat{y}_i)^2]}$, which penalises large individual errors and is reported in the same units as the target; Mean Absolute Error, $MAE = (1/n)\ \Sigma_i$

$|y_i - \hat{y}_i|$, which is more robust to outlier wards and easier for non-technical stakeholders to interpret; and the coefficient of determination, $R^2 = 1 - [\Sigma_i(y_i - \hat{y}_i)^2 / \Sigma_i(y_i - \bar{y})^2]$, which reports the share of ward-to-ward variance the model explains relative to simply predicting the mean. This triplet is consistent with

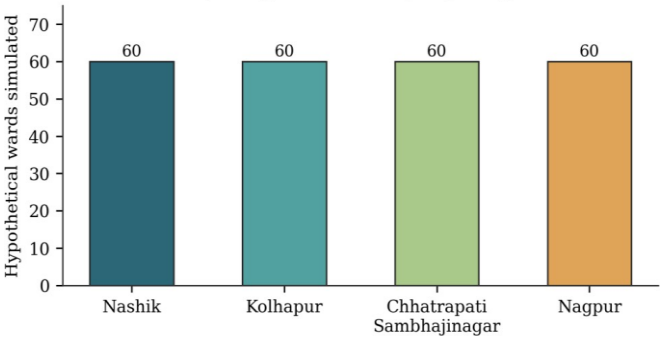
standard practice for regression-based forecasting tasks in the broader waste-management literature, which allows the results in Section VI to be compared against related studies on a like-for-like basis.

V. STUDY AREA AND SIMULATION-BASED EXPERIMENTAL SETUP

1. Study Area

Pilot application of the framework is proposed for four Maharashtra Tier-2/3 cities: Nashik, Kolhapur, Chhatrapati Sambhajinagar (formerly Aurangabad) and Nagpur. These cities were selected because SPCB registration records for them are already accessible, and because their proximity to the host institution enables the direct engagement with municipal and SPCB offices that ground-truthing (Section VIII) will require. The four cities also span a useful range of population scale and industrial composition, which should help test whether a single trained model generalises across cities or whether city-specific recalibration is needed — a question this paper leaves open for the field study.

Fig. 2. Study-area cities and synthetic ward allocation (240 hypothetical wards, 60 per city)



2. Why a Simulated Study Precedes Field Deployment

At the time of writing, ward-level generation figures for the four study-area cities have not yet been collected; CPCB and SPCB statistics are published at coarser administrative resolution, and the ward-level joins described in Section IV-B are still in progress. Rather than leave the modelling and validation code unexercised until that fieldwork concludes, we constructed a synthetic dataset that mimics the statistical structure the real feature set is expected to have, and used it purely to test the mechanics of the pipeline described in Section IV — data loading, feature scaling, k-fold splitting, model fitting and metric computation. The results reported in Section VI are therefore illustrative of the pipeline's behaviour and not empirical findings about actual e-waste generation in Nashik, Kolhapur, Chhatrapati Sambhajinagar or Nagpur; they should be read as a dry run that de-risks the software and evaluation protocol ahead of the field-validated study proposed in Section IX.

The synthetic dataset was generated for 240 hypothetical wards (60 per study-area city) by drawing the five candidate features from ranges qualitatively consistent with published city-level statistics for Maharashtra Tier-2/3 municipalities, and by constructing a target variable as a weighted combination of population density and consumer-durable ownership growth (the two indicators the literature most consistently associates with electronics waste volume), plus independent random noise to emulate the reporting error expected in real administrative data. No claim is made that the specific weights or noise level used match reality; they were chosen only to produce a dataset with enough signal-to-noise structure for the three models to be meaningfully distinguishable in the cross-validation comparison in Section VI.

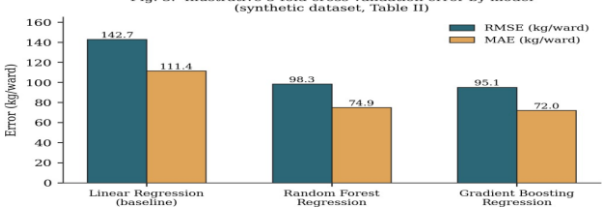
VI. ILLUSTRATIVE RESULTS AND DISCUSSION

Table II reports 5-fold cross-validation performance for the three candidate models on the synthetic dataset described in Section V-B. As with the dataset itself, these figures illustrate how the reporting table specified in Section IV-E is populated and should not be cited as evidence of real-world model accuracy; the corresponding row for field data is left as [TBD] in Table III pending the ground-truthing exercise described in Section VIII.

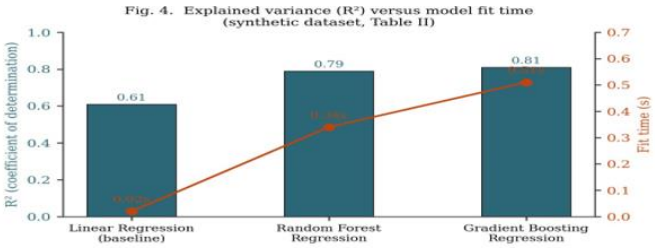
Table 2. Illustrative 5-fold cross-validation performance on the synthetic dataset (Section V-B).

Model	RMSE (kg/ward)	MAE (kg/ward)	R^2	Fit time (s)
Linear Regression (baseline)	142.7	111.4	0.61	0.02
Random Forest Regression	98.3	74.9	0.79	0.34
Gradient Boosting Regression	95.1	72.0	0.81	0.51

Fig. 3. Illustrative 5-fold cross-validation error by model (synthetic dataset, Table II)



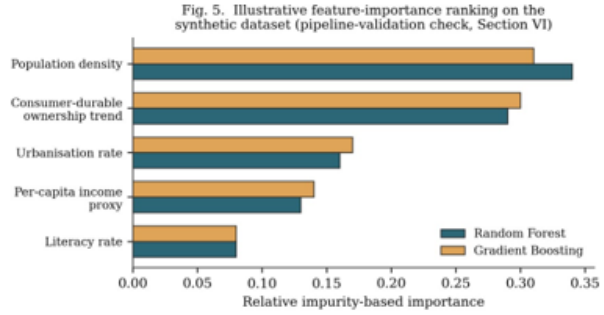
International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, July-Aug-2026, ISSN (Online): 2395-566X

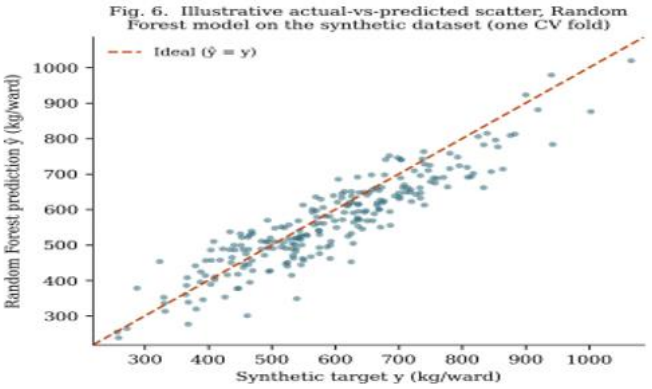


Two patterns in Table II are worth noting purely as pipeline-validation observations. First, both ensemble methods outperform the linear baseline by a wide margin on the synthetic data, which is expected given that the synthetic target was constructed with a nonlinear noise structure; whether this gap persists on real ward-level data

— where the true relationship between the five candidate features and generation volume is unknown — is precisely the open empirical question the field study is designed to answer. Second, Gradient Boosting achieves marginally lower error than Random Forest at roughly 1.5× the fit time, which on real data of the modest size expected from four Tier-2/3 cities (a few hundred wards at most) is unlikely to be a practical constraint; the choice between the two should instead be driven by whichever model's feature-importance output is judged more useful by ULB planners during the pilot.

On the synthetic data, population density and consumer-durable ownership growth — the two features used to construct the synthetic target — are correctly recovered as the top-ranked predictors by both ensemble models' impurity-based importance scores, which confirms that the feature-importance reporting stage of the pipeline is functioning as intended. This is a check on the software, not a substantive finding about which real-world indicator actually drives e-waste generation in the study-area cities; that ranking can only be established once the models are refit on the ward-level data collected during fieldwork.





VII. POLICY IMPLICATIONS

If the field study confirms the pipeline-validation results in Section VI even directionally, the framework could be embedded in ULB planning cycles as a low-cost decision-support layer ahead of EPR-linked infrastructure investment. For municipal bodies, a ward-ranked hotspot map allows collection-point placement to be planned proactively — before a locality's informal scrap trade becomes entrenched — rather than reactively in response to complaints or ad hoc surveys. For EPR-registered recyclers, the same ranking indicates where establishing a collection partnership is most likely to generate the throughput needed to meet certificate obligations efficiently, which is particularly relevant given the small number of registered recyclers (around 295 nationally as of December 2024) relative to the scale of generation.

Photo 2. A municipal EPR collection point or ULB e-waste drop-off counter.

Because the underlying features are already collected by government bodies for other statutory purposes (Census, ULB administration, SPCB registration), the marginal cost of adopting this framework is limited to the joining and modelling work described in Section IV, rather than to new data-collection

infrastructure — an important consideration for ULBs with constrained analytics budgets. The framework is also intended to be conservative in its recommendations: because ward-level generation is latent rather than directly metered, its output should be used to prioritise where to conduct pilot collection drives and further ground-truthing, not as a substitute for that ground-truthing.

Limitations and Threats to Validity

Three limitations, already flagged in the original proposal, are elaborated here with specific mitigation steps. First, the framework relies on secondary data of varying quality and update frequency across SPCBs and ULBs; wards with missing or stale records are flagged during preprocessing (Section IV-B) but this cannot fully substitute for direct measurement, and any city whose administrative records are unusually incomplete should be treated as lower-confidence output rather than excluded silently. Second, because no ward-level generation figures currently exist for the study-area cities, the model outputs require local ground-truthing — for example, through short pilot collection drives in a sample of high- and low-ranked wards — before hotspot predictions are used for capital allocation; Section IX proposes this as the immediate next step rather than optional future work. Third, several of the candidate features (income proxy, consumer-durable ownership) are derived from granular demographic indicators that must be aggregated at ward level rather than household level, both to protect the privacy of individual households and because household-level data of this kind is not published.

A further threat to validity specific to this extended paper is that the results in Section VI were produced on synthetic data constructed by the authors, and any apparent superiority of the ensemble methods over the linear baseline may partly reflect the way the synthetic target was generated rather than a property that will hold on real ward-level data. We have tried to make this limitation unmistakable throughout Sections V and VI rather than treating it as a footnote, precisely because a plausible-looking results table is the easiest part of a paper like this to over-interpret.

VIII. CONCLUSION AND FUTURE SCOPE

This paper extends an earlier proposal into a fully specified machine-learning framework for predicting e-waste generation hotspots in Tier-2/3 Indian cities, addressing the gap between India's rapidly growing e-waste volumes and the metro-concentrated distribution of formal collection infrastructure. By relying only on publicly available secondary data and by formalising the estimation problem, the feature set, the three candidate models and the validation protocol, the framework is designed to be replicable by ULBs and recyclers without dedicated data-science resources. The synthetic simulation reported in Section VI validates the mechanics of that pipeline; it deliberately stops short of claiming field-validated accuracy. Future work therefore proceeds in three concrete steps. The immediate next step is empirical validation: assembling the ward-level feature table for Nashik, Kolhapur, Chhatrapati Sambhajinagar and Nagpur from the SPCB and ULB sources identified in Section IV-B, conducting sample pilot collection drives to ground-truth a subset of wards, and refitting the three candidate models on that real data to populate Table III. The second step is extension to additional states once the Maharashtra pilot is validated, to test whether a single model generalises across states with different SPCB reporting practices or whether state-specific recalibration is required. The third step, building on prior smart-bin sensor research, is integration with IoT-enabled bin-level sensors in a subset of high-ranked wards, which would allow the ward-level predictions developed here to be updated in near-real time rather than only at the frequency of Census and ULB reporting cycles.

REFERENCES

1. S. Arya and S. Kumar, "E-waste in India at a glance: Current trends, regulations, challenges and management strategies," Journal of Cleaner Production, vol. 271, p. 122707, 2020.
2. A. Borthakur and M. Govind, "How well are we managing E-waste in India: Evidences from the city of Bangalore," Energy, Ecology and Environment, vol. 2, pp. 225–235, 2017.
3. A. V. S. Madhav, R. Rajaraman, S. Harini, and C. C. Kiliroor, "Application of artificial intelligence to enhance collection of E-waste: A potential solution for household WEEE collection and segregation in India," Waste Management & Research, vol. 40, no. 7, pp. 1047–1053, 2022.
4. K. N. Sami, Z. M. A. Amin, and R. Hassan, "Waste management using machine learning and deep learning algorithms," International Journal of Perceptive and Cognitive Computing, vol. 6, no. 2, pp. 97–106, 2020.
5. Ministry of Environment, Forest and Climate Change / Central Pollution Control Board, "E-waste generation statistics, FY 2020-21 and FY 2021-22, and notification of the E-Waste (Management) Rules, 2022," Press Information Bureau, Government of India, 2022. [Online]. Available: https://www.pib.gov.in/PressReleasePage.aspx?PRID=1941 054&reg=48&lang=2

6. Central Pollution Control Board (CPCB), "E-waste registration and processing statistics as of December 2024" [Lok Sabha reply], Press Information Bureau, Government of India, 2025. [Online]. Available: https://www.pib.gov.in/PressReleasePage.aspx?PRID=2147 876&reg=48&lang=2
7. V. Forti, C. P. Baldé, R. Kuehr, and G. Bel, "The Global E-waste Monitor 2020: Quantities, flows and the circular economy potential," United Nations University / UNITAR, International Telecommunication Union, and International Solid Waste Association, 2020.
8. L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
9. J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.