A Serverless Intelligent Framework for Privacy-Preserving Municipal Solid Waste Classification and Circular Economy Assessment: Empirical Evidence from Deolali Pravara

Chhaya S Galande, Nitin E.Kakade

Abstract: *The digital transformation of municipal solid waste management in resource-constrained regions continues to face persistent challenges related to infrastructure costs, technical capacity limitations, and data confidentiality concerns. This investigation presents an integrated browser-native intelligent system designed for real-time waste categorization and generation forecasting using a zero-infrastructure computational approach. Following a comprehensive 7-phase design science methodology, the research evaluates 13 classification models and 12 forecasting algorithms using 9,839 historical records collected from the Deolali Pravara municipal jurisdiction spanning 2022-2024. Empirical findings reveal that the Random Forest ensemble method attains a classification accuracy of 93.1%, while Support Vector Regression and Linear Regression demonstrate robust predictive stability with an* R^2 *coefficient of 0.929. The framework additionally quantifies circular economy potential, estimating a material recovery value of ₹13,046,534. By executing all computational operations entirely on client-side infrastructure, this solution ensures complete data sovereignty with negligible operational expenditures, offering a scalable pathway for data-informed waste governance in developing urban contexts.*

Chhaya S Galande (csgalande27@gmail.com) Head, Department of Computer Science, Padmashri Vikhe Patil College of Arts, Science and Commerce, Pravaranagar.
Nitin E.Kakade (nitin.kakade@pravara.in) Department of Computer Science, Padmashri Vikhe Patil College of Arts, Science and Commerce, Pravaranagar.

ISSN: 0262-8104 (print) 1756-3488 (online)
Waterlines Vol. 44 No. 1s

Introduction

1.1 The Escalating Global Waste Crisis

Municipal solid waste generation constitutes one of the most pressing environmental challenges confronting contemporary urban societies. Current global waste production stands at approximately 2.01 billion tons annually, with projections indicating a substantial escalation to 3.40 billion tons by 2050 if current consumption patterns persist. This trajectory reflects complex interplays between population growth, urbanization, economic development, and changing consumption behaviors that collectively drive waste generation rates upward across all regions.

The environmental ramifications of improper waste disposal extend far beyond visual pollution and land degradation. Decomposing organic matter in landfills releases substantial quantities of methane—a greenhouse gas with a global warming potential 28-36 times greater than carbon dioxide over a 100-year timeframe. Municipal solid waste disposal sites currently contribute approximately 11% of global anthropogenic methane emissions, representing a significant yet often overlooked contributor to climate change. This methane generation occurs primarily through anaerobic decomposition of organic fractions, which constitute a substantial portion of municipal waste streams in developing economies.

1.2 The Indian Urban Waste Scenario

India's urban centers present a particularly acute manifestation of the global waste challenge. Cities generate over 70% of the nation's total waste output, yet only 20-25% undergoes any form of formal processing or treatment. This processing gap reflects multiple systemic challenges including inadequate infrastructure, limited financial resources, institutional capacity constraints, and technological deficiencies that collectively undermine effective waste management outcomes.

The informal sector plays a substantial role in India's waste management ecosystem, with waste pickers and itinerant buyers recovering recyclable materials from disposal sites and urban neighborhoods. While this informal recovery system contributes meaningfully to resource conservation and livelihood generation, it operates largely outside formal governance structures, resulting in significant health hazards, unsafe working conditions, and environmental externalities that remain unaddressed by regulatory frameworks.

1.3 The Analytical Gap in Smaller Municipalities

Large metropolitan areas in India have increasingly adopted server-based artificial intelligence platforms for waste analytics, leveraging cloud computing resources and specialized technical teams to enhance operational efficiency. However, small and medium-sized urban local bodies face what this study characterizes as an "analytical gap" across multiple dimensions:

Capacity Limitations: Smaller municipalities typically lack integrated analytical platforms capable of simultaneously addressing classification, forecasting, and conversion analysis requirements. This fragmentation impedes holistic decision-making and prevents systematic optimization of waste management operations.

Resource Constraints: The acquisition and maintenance of server infrastructure, software licensing, and specialized technical personnel represent substantial financial burdens that exceed the budgetary capacities of smaller municipalities. Cloud-based solutions, while reducing upfront capital expenditures, introduce ongoing operational costs that remain prohibitive for resource-constrained local governments.

Data Privacy Concerns: The transmission of municipal waste data to external servers raises legitimate privacy and security concerns. Local authorities may be subject to data protection regulations, citizen privacy expectations, and security requirements that complicate the adoption of external platforms. The potential for data breaches or unauthorized access to sensitive municipal information creates institutional hesitation toward cloud-based solutions.

1.4 esearch Objectives and Contributions

This investigation pursues four interconnected objectives designed to address the analytical gap identified above:

1. Characterization of Municipal Waste Patterns: Through systematic analysis of historical waste data using data science techniques, the study aims to understand waste generation characteristics, temporal patterns, and compositional variations across different urban zones.
2. Resource Recovery Opportunity Assessment: By examining waste-to-product conversion pathways, the research identifies economically viable recovery opportunities that can generate revenue while reducing landfill burden.

3. Machine Learning Benchmarking: The systematic evaluation of 25 distinct algorithms establishes performance benchmarks specific to municipal waste classification and forecasting tasks, providing guidance for future implementations.
4. Client-Side Architectural Development: The creation of a functional, browser-native module demonstrates the feasibility of serverless waste analytics while maintaining data privacy and computational efficiency.

The primary contribution of this research lies in demonstrating that sophisticated analytical capabilities can be delivered through zero-infrastructure, privacy-preserving architectures suitable for resource-constrained municipal contexts.

Literature Review

2.1 Foundational Research in Waste Characterization

The systematic characterization of municipal solid waste composition has long been recognized as essential for effective management planning. Early studies established fundamental frameworks for waste categorization based on material types, moisture content, calorific value, and biodegradability. These foundational classification schemes have evolved to incorporate increasingly sophisticated analytical dimensions including heavy metal concentrations, organic contaminant levels, and potential for various recovery pathways.

Garcia and colleagues (2005) investigated the utilization of organic waste fractions as animal feed, identifying significant nutritional potential while simultaneously recognizing critical safety constraints. Their research demonstrated that while organic municipal waste contains substantial nutrient density suitable for animal consumption, proper heat treatment remains essential to eliminate pathogenic organisms and meet food safety standards. This work highlighted the complex trade-offs inherent in waste valorization approaches, where resource recovery potential must be balanced against health and safety considerations.

2.2 Energy Recovery and Environmental Impact Assessment

The conversion of municipal waste to energy has received substantial research attention as a dual-purpose solution addressing both waste disposal and renewable energy generation challenges. Kaur and associates (2023) examined biomethanation and gasification technologies as mechanisms for reducing greenhouse gas emissions while decreasing fossil fuel demand. Their analysis indicated that controlled anaerobic digestion of organic waste fractions not only captures methane for energy

utilization but also prevents its uncontrolled release from landfills, yielding net environmental benefits substantially exceeding simple waste disposal approaches.
The environmental impact assessment literature has increasingly employed life cycle analysis methodologies to quantify the comprehensive environmental consequences of different waste management approaches. These studies typically incorporate multiple impact categories including global warming potential, acidification, eutrophication, and resource depletion, providing a more nuanced understanding of environmental trade-offs than single-metric assessments.

2.3 Forecasting Approaches in Waste Management

Prediction of waste generation patterns has emerged as a critical capability for operational planning, infrastructure sizing, and resource allocation. Rathod and Patel (2025) conducted a comprehensive review of forecasting models applied to municipal solid waste, finding that over 30% of recent studies utilized Artificial Neural Networks due to their capacity to capture complex nonlinear relationships among urban variables. Their review identified socioeconomic factors, seasonal patterns, and demographic characteristics as the most influential predictors in waste generation models.

Traditional statistical approaches including autoregressive integrated moving average models and exponential smoothing methods continue to find application in waste forecasting, particularly in contexts with limited data availability or where interpretability is prioritized over predictive accuracy. The choice between machine learning and statistical approaches typically involves balancing predictive performance against computational requirements and model interpretability considerations.

2.4 Real-Time Segregation Technologies

The development of automated segregation technologies has progressed significantly, enabling real-time material identification without manual sorting. Pawar and colleagues (2013) introduced the "SMART SORT" system utilizing audio classification techniques to distinguish materials based on acoustic signatures. Their system achieved 94.4% accuracy for materials including glass and plastic, demonstrating the viability of non-visual sensing approaches for automated material identification.

Subsequent research has explored optical sorting, near-infrared spectroscopy, and artificial vision systems for real-time waste categorization. These technologies enable automated material recovery facilities to achieve higher purity outputs while reducing labor costs and worker exposure to hazardous materials. However, the capital intensity of such systems has limited their adoption primarily to large-scale facilities in developed economies.

2.5 Research Gap and Contribution

While the existing literature contains valuable contributions across specialized domains of waste management, limited research has addressed the integration of classification, forecasting, and valuation functions within a single accessible platform suitable for small-to-medium municipalities. Furthermore, the privacy and infrastructure challenges inherent in cloud-based solutions have received insufficient attention in the waste analytics literature. This research addresses these gaps by providing an integrated, privacy-preserving, serverless framework that delivers comprehensive analytical capabilities to resource-constrained local governments.

Research Methodology

3.1 Design Science Research Framework

This investigation adopts Design Science Research methodology, a problem-solving paradigm particularly suited to creating and evaluating innovative artifacts designed to address practical organizational challenges. The DSR approach is characterized by its emphasis on utility, iterative development, and rigorous evaluation—characteristics well-aligned with the goals of developing a functional waste analytics system.

The 7-phase implementation process followed in this research comprises:
Phase 1 - Problem Definition: Comprehensive stakeholder consultations to identify requirements, constraints, and expectations for waste analytics in Deolali Pravara. This phase involved engagement with municipal officials, waste management personnel, and community representatives to understand operational realities and information needs.

Phase 2 - Data Collection: Systematic sourcing of historical waste records from municipal databases spanning 2022-2024, resulting in 9,839 individual observations. Data collection procedures ensured consistency with municipal record-keeping practices while verifying data completeness and integrity.

Phase 3 - System Design: Development of a 3-tier client-side architecture incorporating data management, analytical computation, and visualization components within a unified browser-native environment.

Phase 4 - Processing Pipeline: Construction of an Upload-Edit-Clean workflow enabling data ingestion, quality assessment, and preprocessing operations essential for reliable analysis.

Phase 5 - Machine Learning Module: Implementation of 25 algorithms encompassing classification and forecasting approaches appropriate for municipal waste applications.

Phase 6 - Analytics and Visualization: Design of interactive dashboards supporting exploratory analysis, performance monitoring, and economic valuation.

Phase 7 - Validation: Systematic evaluation of system outputs using established metrics including Root Mean Square Error, Mean Absolute Percentage Error, and classification accuracy.

Study Area Description

Deolali Pravara, the study location, is a municipal town situated in Maharashtra State, India. With a population of approximately 30,4 residents, the municipality represents a typical small-to-medium Indian urban center characterized by mixed residential, commercial, and institutional land uses. The town's waste management system operates under the administrative jurisdiction of the municipal council, with waste collection, transportation, and disposal responsibilities distributed across 18 geographic wards.

The urban structure of Deolali Pravara encompasses diverse land use patterns including densely populated residential neighborhoods, commercial districts, educational institutions, and agricultural peri-urban zones. This diversity contributes to heterogeneous waste generation patterns across the municipal area, with compositional variation reflecting the socioeconomic characteristics and functional specializations of different zones.

3.3 Data Model and Attributes

The waste data structure follows a standardized format designed to capture essential attributes for analysis:

Temporal Dimension: Collection dates recorded in ISO 8601 format enable temporal analysis of generation patterns, seasonal variation assessment, and time-series forecasting applications. The 2022-2024 timeframe provides sufficient historical depth for identifying multi-year trends while maintaining data quality and consistency.

Spatial Dimension: Zone identifiers corresponding to 18 municipal wards facilitate spatial analysis, enabling identification of geographic patterns, zone-specific generation characteristics, and infrastructure planning implications. This spatial granularity supports targeted interventions and resource allocation.

Compositional Dimension: Waste_Type categorization into Plastic, Organic, Paper, Metal, and Glass classes enables material-specific analysis essential for recovery planning. These categories align with conventional waste classification frameworks while maintaining practical utility for municipal operations.

Quantitative Dimension: Quantity_kg numeric values provide the basis for generation rate calculations, trend analysis, and conversion valuation. The use of standardized mass units facilitates comparison across zones, time periods, and material categories.

3.4 Data Preprocessing Methodology

The preprocessing pipeline is designed to ensure analysis integrity through systematic identification and treatment of data quality issues:

Duplicate Record Elimination: A composite key incorporating Date, Zone, and Waste_Type attributes enables precise identification of redundant records. The elimination algorithm operates with O(n) complexity, ensuring efficient processing of the 9,839-record dataset. Duplicate records most frequently arise from data entry errors, redundant reporting procedures, or system synchronization issues within the municipal data collection infrastructure.

Outlier Detection Protocol: The Interquartile Range methodology is employed for statistical outlier identification, with boundaries defined as:

- IQR = Q3 - Q1 (where Q3 and Q1 represent the third and first quartiles respectively)
- Upper Bound = Q3 + (1.5 × IQR)

- Observations exceeding the Upper Bound are flagged as high-quantity outliers

The IQR approach is selected for its robustness to extreme values and its non-parametric nature, making it suitable for waste quantity distributions that typically exhibit right-skewed characteristics. Outlier identification triggers investigation rather than automatic removal, allowing subject matter expertise to guide treatment decisions.

Machine Learning Implementation

4.1 Classification Algorithm Suite

Thirteen distinct classification algorithms were implemented and evaluated for waste category prediction:

K-Nearest Neighbors (KNN): This instance-based learning algorithm employs Euclidean distance calculations for similarity assessment, with the formulation $d(x, xi) = \sqrt{\Sigma(xj - xij)^2}$. The algorithm classifies new instances by majority vote among k nearest neighbors in feature space. KNN's non-parametric nature and intuitive operation make it particularly suitable for applications requiring interpretability, though performance may degrade with high-dimensional data.

Random Forest (RF): An ensemble learning method constructing 100 decision trees with Gini Impurity as the splitting criterion. Random Forest operates through the principle of "wisdom of crowds," combining predictions from multiple decorrelated trees to achieve superior generalization compared to individual decision trees. The algorithm's robustness to overfitting, capability for handling mixed data types, and inherent feature importance assessment make it particularly well-suited to waste classification applications.

Logistic Regression: Multi-class logistic regression employs the Softmax function for probability estimation:

$P(y = k|x) = \exp(w\_k^T x) / \Sigma \exp(w\_j^T x)$

This probabilistic approach provides both class predictions and confidence estimates, supporting informed decision-making in waste management operations. The linear decision boundary of logistic regression offers advantages in terms of computational efficiency and interpretability, though it may struggle with complex nonlinear class boundaries.

Support Vector Machine (SVM): Employing Radial Basis Function kernels to project data into higher-dimensional spaces where linear separation becomes feasible. SVM optimization seeks to maximize the margin between class boundaries, providing robustness to outliers and generalization capability. The kernel trick enables SVM to capture complex nonlinear relationships while maintaining computational efficiency.

4.2 Forecasting Algorithm Implementation

Five primary forecasting models were implemented to capture distinct characteristics of waste generation patterns:

Linear Regression: This fundamental approach models the relationship between time and generation as $y = \beta_0 + \beta_1 t + \varepsilon$, providing baseline trend identification and simple forecasting. The linear assumption enables straightforward interpretation of direction and magnitude of temporal trends, though it may inadequately capture seasonal patterns and nonlinear dynamics.

Prophet (Facebook): A decomposition-based forecasting approach separating time series into trend $g(t)$, seasonality $s(t)$, holiday effects $h(t)$, and error ε components: $y(t) = g(t) + s(t) + h(t) + \varepsilon$. Prophet's capacity for handling missing data, robust trend change detection, and user-friendly parameterization make it particularly valuable for municipal waste forecasting where data quality issues and human-specified holidays influence generation patterns.

Long Short-Term Memory (LSTM): A specialized recurrent neural network architecture designed to capture long-term dependencies through the operation of forget, input, and output gating mechanisms. LSTM's capacity for learning from sequence data without arbitrary window length constraints enables capture of complex temporal patterns that may extend across multiple seasonal cycles. The architecture's representational power comes at the cost of increased data requirements, training complexity, and computational demands.

Support Vector Regression (SVR): An adaptation of SVM principles to regression tasks, SVR seeks to find a function that deviates from actual targets by no more than a specified margin while remaining as flat as possible. The use of kernel functions enables capture of nonlinear relationships while maintaining the regularization properties characteristic of SVM approaches.

Gradient Boosting: An ensemble method constructing predictive models sequentially, with each subsequent model focusing on correcting errors of its predecessors. Gradient boosting demonstrates particular effectiveness in capturing subtle patterns and interactions while providing feature importance insights.

5. System Architecture

5.1 Client-Side Philosophy and Design Principles

The system architecture embodies a fundamental departure from conventional server-based approaches through its commitment to complete client-side execution. This design philosophy reflects the recognition that infrastructure constraints and privacy concerns represent primary barriers to municipal waste analytics adoption.
Zero-Backend Implementation: Leveraging Next.js 14 and TypeScript, the system operates entirely within the browser environment, eliminating server requirements for computation, storage, or application logic. This approach reduces operational costs to near-zero while simplifying deployment and maintenance.

Data Sovereignty: By performing all processing on client machines, the architecture ensures that municipal waste data never leaves the user's control. This eliminates concerns regarding server breaches, unauthorized access, or third-party data exposure that commonly arise in cloud-based analytics solutions. The privacy-preserving nature of client-side processing addresses a fundamental barrier to adoption identified in stakeholder consultations.

Persistent Storage: LocalStorage and IndexedDB provide browser-native storage mechanisms for data persistence across sessions. These client-side storage options enable retention of uploaded data and analysis results without requiring network connectivity or external infrastructure. The capacity of IndexedDB for storing substantial datasets ensures the system can handle municipal archives of significant size.

Visualization Performance: Recharts library implementation achieves 60 FPS interactive rendering for datasets up to 10,000 records, enabling responsive exploration and analysis. The visualization component supports multiple chart types, user interactions, and dynamic filtering to facilitate exploratory data analysis and pattern identification.

5.2 Data Processing Workflow

The system implements a structured data processing workflow ensuring analytical integrity:

Upload Stage: Users upload municipal waste data in standard formats including CSV and Excel files. The system validates file structure, data types, and required fields to ensure compatibility with subsequent processing stages.

Edit Stage: A user interface presents the uploaded data for review and manual correction. Users may edit individual records, add annotations, and record metadata regarding data sources and collection procedures.

Clean Stage: Automated preprocessing operations are applied to identify duplicates, detect outliers, and flag data quality issues. The user receives a quality report summarizing identified issues and recommended actions.

Analysis Stage: Machine learning models are trained on the prepared data, with results presented through interactive visualizations and summary metrics. Users can compare algorithm performance, explore predictions, and generate economic valuations.

Results and Discussion

6.1 Data Quality Assessment

The preprocessing pipeline demonstrated efficient and effective processing of the raw dataset. Processing completion was achieved in 4.2 seconds, enabling rapid data preparation despite the 9,839-record dataset size. The system identified 372 flagged issues requiring user attention, primarily comprising outliers in quantity measurements and potential data entry inconsistencies. Following user review and resolution of flagged issues, the system achieved a final data quality score of 99.2%, indicating high reliability for subsequent analytical operations.

The data quality assessment revealed several notable patterns worth consideration for future implementation: outlier concentrations in specific zones suggested potential data collection protocol variations, while temporal patterns in flagged issues indicated potential seasonal effects on data quality. These findings inform recommendations for data collection process improvements.

6.2 Classification Algorithm Performance

The comparative evaluation of classification algorithms revealed meaningful performance differentiation across the 13 implemented approaches:

Random Forest attained the highest overall accuracy at 93.1%, confirming the effectiveness of ensemble methods for waste category classification. The combination of bagging and feature randomness in Random Forest provided robust generalization while minimizing overfitting risk. Analysis of feature importance within the Random Forest model identified quantity and zone as the most influential predictors, with temporal features providing secondary explanatory power.

Support Vector Machine demonstrated the highest precision at 93.6%, indicating exceptional capability in minimizing false positives. This characteristic makes SVM particularly valuable for applications where false classification of non-recyclable materials as recyclable may lead to contamination of recovery streams. The precision advantage of SVM must be weighed against its lower recall in material-specific contexts.

LightGBM achieved the best recall at 95.0%, indicating superior capability in identifying true positives for each waste category. This characteristic supports applications where failing to identify recyclable materials represents a greater concern than occasional false positives. The recall advantage of LightGBM comes at some cost to precision, requiring appropriate weighting based on operational priorities.

6.3 Forecasting Model Evaluation

The forecasting evaluation revealed the relative strengths of different modeling approaches:

Support Vector Regression and Linear Regression achieved an R^2 coefficient of 0.929, indicating that these approaches explain over 92% of variance in generation patterns. The strong performance of both approaches suggests that waste generation in Deolali Pravara follows largely linear trends with limited nonlinear dynamics requiring complex modeling. The close correspondence between SVR and linear regression performance may reflect the absence of substantial nonlinear patterns in the temporal data.

Gradient Boosting provided the best MAPE at 6.6%, indicating the lowest percentage deviation between predictions and actual values. The superior MAPE performance of Gradient Boosting suggests capacity to capture subtle patterns that linear approaches may overlook, despite similar R^2 performance. The selection between linear approaches and Gradient Boosting for forecasting applications depends on whether interpretability or predictive accuracy represents the higher priority.

6.4 Economic Valuation and Circular Economy Potential

The conversion module identified 13 distinct product pathways for material recovery, with substantial variation in economic value:

Metal Recovery: Metal waste contributed ₹8,506,183, representing 65.2% of total estimated value. The high value reflects both the substantial quantities of metal waste generated and the favorable market prices for recovered metals. Metal recovery represents the most economically attractive pathway, supporting prioritization of metal recovery infrastructure development.

Plastic Valorization: The potential value of plastic granule production was estimated at ₹2.1 million. While lower than metal recovery, plastic valorization represents a meaningful revenue stream that can contribute to waste management sustainability. The relatively lower value of plastic recovery reflects market conditions and contamination challenges affecting plastic quality and marketability.

Environmental Impact Assessment: The framework estimates that diversion of organic waste from landfills would avoid 429.5 tonnes of CO_2e emissions through methane capture avoidance. This substantial environmental benefit provides additional justification for organic waste recovery investments beyond direct economic returns.

Conclusion and Future Directions

This research successfully addresses the "analytical gap" facing small and medium-sized municipalities by providing a high-performance, cost-free, and privacy-preserving tool for municipal waste management. The demonstration that sophisticated machine learning algorithms including Random Forest and SVR can achieve near-state-of-the-art results on local browser hardware establishes the technical feasibility of serverless waste analytics.

The practical contributions of this research extend beyond technical performance to encompass institutional and economic dimensions. The quantification of circular economy potential provides municipal decision-makers with actionable information for resource recovery investments. The privacy-preserving architecture addresses institutional concerns regarding data sovereignty that have hindered adoption of cloud-based solutions.

References

Garcia, A.J., Esteban, M.B., Márquez, M.C., & Ramos, P. (2005). Biodegradable municipal solid waste: Characterization and potential use as animal feedstuffs. Waste Management, 25(8), 780-787.

Kaur, A., Singh, P., & Sharma, R. (2023). Energy recovery from municipal solid waste through biomethanation and gasification: A review. Materials Today Proceedings, 81(6), 234-241.

Rathod, A.R., & Patel, V.M. (2025). Forecasting models for municipal solid waste generation: A systematic review. International Journal of Scientific Research and Development, 10(4), 112-119.

Pawar, A.S., Shinde, S.K., & Patil, S.N. (2013). SMART SORT: Audio-based classification for solid waste segregation. International Journal of Current Research and Technology, 12(05), 789-796.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.

Yelwande, C.S. (2025). Study and Evaluation of Municipal Solid Waste with Handling Approaches Using Data Science Techniques. Doctoral dissertation, Shri JIT University, Rajasthan, India.