Journal of Engineering and Technological Sciences

Research Paper

A Comparative Study of Advanced Machine Learning and Deep Learning Models for Municipal Solid Waste Forecasting: A Case Study of Surat, India

Abhijit R. Rathod[1,*], & Vinodkumar M. Patel[2]

[1]Gujarat Technological University, GTU Campus, Nr. Visat Three Road, Visat-Gandhinagar Road, Chandkheda, Ahmedabad-382424, Gujarat, India
[2]Shantilal Shah Engineering College, New Sidsar Campus, Vartej, Sidsar, Bhavnagar - 364060, Gujarat, India

*Corresponding author: arrathod1709@gmail.com

Abstract

To make the city planning sustainable, especially in rapidly growing cities of the world like Surat in India, the implementation of effective waste management is crucial. The primary factor governing this is the ability to accurately predict the quantity of municipal solid waste (MSW) likely to be generated. This study presents a comprehensive comparative analysis of various predictive models including linear regressions, kernel approaches, gradient boosting as well as the deep learning architectures. Using historical data from Surat, systematic preprocessing and feature engineering generated 419 features representing temporal, socio-economic, climatic, COVID-19, and mobility factors. The novel contribution of this study is the systematic feature engineering framework that explicitly encodes temporal structure (419 engineered features including lagged values, rolling statistics, and seasonal decomposition), enabling simple linear models to capture complex waste generation patterns. Ten distinct models, ranging from statistical approaches to machine learning and deep learning were evaluated and compared. Advanced ensemble models, including LightGBM (R^2 = 0.983), CatBoost (R^2 = 0.977), and XGBoost (R^2 = 0.970) demonstrated strong performance. The best-performing models (OLS and Gaussian Process Regression) achieved R^2 = 0.997 with Mean Absolute Percentage Error (MAPE) = 1.43%. In this study, linear models trained within a few milliseconds and achieved per-sample inference times on the order of 0.004-0.008 ms, whereas the tuned MLP and tree-based ensembles required seconds of training and millisecond-level inference, corresponding to differences of roughly two to three orders of magnitude in computational cost. Other notable performers include Lasso regression (R^2 = 0.979), tuned MLP (R^2 = 0.968), and Random Forest (R^2 = 0.960). The results demonstrate that feature engineering has greater influence on forecasting accuracy than model complexity, with OLS using engineered features (R^2 = 0.997) outperforming the MLP model (R^2 = 0.966) by approximately 3.1% while providing substantially faster predictions. Feature importance analysis identified lagged MSW values, rolling statistics, demographic indicators, festival effects, and COVID-19 lockdown impact as key predictors. The research finds that adoption of systematically designed feature engineering framework is a valuable tool for MSW management. The study provides comprehensive model benchmarking and practical recommendations for policymakers pursuing sustainable urban development.

Keywords: *deep learning; forecasting; gradient boosting; machine learning; municipal solid waste; Surat; time series.*

Introduction

The 21st century is witnessing unprecedented and rapid urbanization across the globe. The global economy is being reshaped by the demographic shift and dramatic increase in generation and complexity of municipal solid waste (MSW). These rapidly growing cities face huge pressure to handle their waste management systems as a result of this change. What if city management fails to plan adequately and accurately? The consequences are severe: overflowing landfills create public health hazards; encourage illegal dumping that pollutes water sources and soil; and cause transportation disruptions that paralyze municipal operations. For example; Surat; India; one of India's fastest-growing cities saw waste generation surge from 3;800 tonnes per day in 2009 to 6;200 tonnes per day in 2024. Without accurate forecasting; municipal authorities struggle to deploy collection vehicles; schedule landfill operations; and allocate limited budgets effectively.

ISSN: 2337-5779
J. Eng. Technol. Sci. Vol. 58, No. 5, 2026, 613-629
DOI: 10.5614/j.eng.technol.sci.2026.58.5.1

To cope with these issues; innovative and data-driven solutions are the need of the hour (World Bank Group, 2022). By the year 2050; the world's urban population is estimated to cross the 6 billion marks; meaning it will nearly double compare to the urban population in 2000. This will lead to a huge increase in the amount of MSW produced and will create new problems for city infrastructure and the environment overall (UN-Habitat, 2018). Urban migration causes populations to become increasingly dense. We are also witnessing the change in the consumption patterns. The waste streams are becoming more and more diverse with more plastics; packaging; and electronic waste (Y.-C. Chen, 2018). The transition from rural to urban lifestyles causes the generation of non-biodegradable waste; putting a huge pressure on current collection and disposal systems (Chatterjee, 2024). When cities grow quickly; the infrastructure must match the pace to keep their management of waste up to date. Often cities' administration fails to do so; which can further result in issues such as overflowing landfills; illegal dumping; and damage to the environment (Smyth, 2024; Zhang et al., 2024). In developing countries; these problems are made worse by a lack of funding; inadequate rules and regulations; and a lack of technical skills (Smyth, 2024). MSW management has emerged as one of the critical challenges in modern times; particularly in most parts of the world which are observing urbanization taking place at the speed of light. Naturally; expanding cities with growing populations face increased volumes and complexity of waste generated. Ultimately; this is putting immense pressure on urban infrastructure and environmental resources available. We cannot stop this enormous and rapid growth of cities; but we can at least make an attempt to predict how much MSW is likely to be generated in the next couple of days; months; or years; which facilitates more effective future planning and by this way protecting the public health and environment also. This is precisely why the ability to accurately predict MSW generation in rapidly growing cities has never been more critical. It plays a pivotal role in reaching goals for the sustainable city management. From a practical perspective; the dilemma faced by municipal administrators is: what if they overestimate waste generation? This results in wasted budget resources on unnecessary infrastructure expansions. And what if they underestimate? Underestimation results in system collapse; public anger; and irreparable environmental damage. Whereas accurate forecasting directly translates to the cost saving (approximately 15- 25% reduction in operational costs by applying optimized vehicle routing) and protecting the environment. That is the reason why MSW forecasting is not merely an academic exercise but is; in actuality; a critical operational requirement for the rapidly growing cities of the world; especially in developing economies having a limited budget and cannot afford any financial mismanagement.

To develop a robust model capable of accurately predicting MSW; the model's ability to capture various waste generation patterns is carefully studied and verified. These patterns are not as simple as they appear in growing cities; given their dependencies on multiple interacting factors like changes in demography (population growth; migration patterns; etc.); economic activities (industrial output; consumption levels; etc.); seasonal variations (effect of monsoon; peak observed in holiday consumption; etc.) and unexpected disruptions (like COVID-19 lockdowns; festivals causing temporary population surges; etc.). All these factors interact non-linearly; making forecasting inaccurate if only historical patterns are considered. The model must incorporate explicit feature engineering.

The conventional approach to predictive modelling fails to capture the complex; nonlinear relationships among variables significantly affecting MSW generation when proper feature engineering is not applied during preprocessing. Despite noticeable advances in machine learning and waste management research; critical gaps still exist in MSW forecasting literature. First; existing studies often give more importance to the architectural novelty (like ensemble combinations; applications of transformations; etc.) without systematically evaluating the effect of simple feature engineering approach; which could deliver equivalent or superior performance. This major gap is directly addressed in our study through the comprehensive comparative analysis we've performed.

Second; waste forecasting literature concentrates heavily on developed-economy urban contexts; empirically-grounded forecasting models specific to developing-economy high-growth cities remain scarce; despite distinct waste generation drivers in these contexts (informal waste economy; rapid population changes; infrastructure constraints). Third; prior studies typically report single or dual evaluation metrics (R^2; RMSE); multidimensional performance assessment addressing scale-dependence; directional accuracy; and computational efficiency is largely absent. Fourth; characterization of computational efficiency; which is essential for deployment feasibility in resource-constrained municipalities receives minimal attention in waste forecasting literature. This study addresses each of these gaps: (1) systematic model comparison across ten approaches; from simple to complex; (2) developing-economy context (Surat; India); with findings transferable to comparable cities; (3) eight complementary evaluation metrics; (4) explicit computational profiling enabling deployment feasibility assessment.

This deficiency ultimately leads to the suboptimal allocation of resources by municipal authorities. Recent studies have suggested the use of machine learning and deep learning approaches for handling non-linear; multi-dimensional feature spaces that are also temporal dependent. Recent studies have appreciated the usage of various machine learning (ML) methodologies like support vector machines; decision trees; and ensemble methods; which exhibit better predictive

DOI: 10.5614/j.eng.technol.sci.2026.58.5.1

accuracy and are more robust than the traditional methods. A substantial scope for improvement remains; owing to gaps in integrating diverse data sources; insufficient attention to temporal dynamics; and limited assessment of model generalisability across different urban settings. This study addresses these challenges by developing and evaluating a comprehensive set of advanced forecasting models for monthly MSW generation in the city of Surat; India. Surat; being India's 9th largest city with rapid population growth; was chosen as the case study for this study. Surat's waste management challenges reflect those faced by many cities of developing nations worldwide. Surat city's waste generation patterns show clear seasonal variations (impacts of festivals; monsoons; etc.) and disruptions (caused by COVID-19 lockdowns) make it an ideal choice for developing forecasting models able to handle complex and real-world conditions. Therefore; it is assumed that solutions developed for the city of Surat can be directly transferable to other Indian and South Asian cities that are urbanizing rapidly and facing similar challenges. To address these challenges; this study poses three fundamental questions that municipal practitioners face when designing forecasting systems:

1. How do advanced machine learning models compare in their accuracy for predicting MSW generation at the city level?
2. Which features and data transformations improve the predictive performance of the developed models?
3. Do these models perform reliably under sudden disruptions (e.g.; festivals; seasonal variations) and gradual societal shifts such as those caused by COVID-19 lockdowns?

Recent advances in machine learning have expanded the methodological toolkit for time-series analysis across various domains worldwide. Attention mechanisms; which dynamically weight feature importance; have proved to highly effective in capturing complex temporal dependencies (J. Chen, Fan, et al., 2025). Similarly; the graph-based neural networks have demonstrated the capability in modelling non-linear relationships in multivariate data (J. Chen et al., 2026). Noticeably; the advanced optimization strategies such as hyperparameter self-optimization have improved model performance across diverse classification tasks (J. Chen, Cui, et al., 2025). Meanwhile; nonlinear feature decomposition combined with deep temporal-spatial learning has been shown to significantly enhance pattern recognition in sequential bio signal data; as demonstrated in recent deep temporal- spatial learning work (J. Chen, Fan, et al., 2025) and in single-channel sEMG-based lower limb motion recognition studies in the IEEE Sensors Journal (Wei et al., 2026). While these innovations originate primarily from signal processing domains; they reflect broader trends in machine learning that warrant consideration in infrastructure forecasting applications.

Following are the four distinct methodological contributions this study makes to the MSW forecasting which clearly differentiate this research from the existing approaches observed in current literature:

1. Empirical Evidence Generation: Instead of generating a novel architecture our study provides a systematic empirical comparison of ten established models ranging from OLS to deep learning. We have applied a comprehensive feature engineering to hep practioners select models based on empirical evidences which is essential for decision support.
2. Systematic Feature Engineering Framework: The research proposed explicitly articulates 419 feature engineering pipeline incorporating temporal decomposition (STL); lagged variables (1-12 months); rolling statistics (3;6; 12 month window) and domain specific indicators (like monsoon; COVID-19; festivals; population growth; etc.) This approach represents a methodological contribution which can be easily transferable across municipal forecasting applications.
3. Multidimensional Evaluation Framework: To establish evaluation standards for future waste forecasting research; instead of depending upon single metric evaluation we've integrated eight complementary metrics like R^2; MAE; RMSE; MAPE; coefficient of variation; Theil's U; Mean Directional Accuracy and Residual Standard Deviation.
4. Computational Efficiency Characterization: In contemporary waste forecasting literature systematic documentation of model deployment (training time; inference speed; parameter count; model size) is typically not available. The trends prevailing currently emphasizes architectural innovation over systematic adoption of feature engineering; employs limited evaluation metrics and often neglects characterization of computational efficiency of the model developed. Our study adopts practical model deployment considerations; prioritizing empirical comparison; feature engineering quality provides municipal administrators and policymakers actionable guidance essential for the design of forecasting system in resource limited developing economy contexts.

The remainder of this paper is organized as follows. Section 2 describes the data corpus and expanded feature space; including the preprocessing and feature engineering pipeline. Section 3 presents the forecasting model architectures; covering linear models; ensemble methods; and deep learning approaches. Section 4 reports the empirical results and comparative analysis of model performance and computational efficiency. Section 5 discusses the methodological contributions; limitations; and implications for municipal waste management practice; and Section 6 concludes the study and outlines directions for future research.

Methodology

Data Corpus and Expanded Feature Space

This study primarily focuses on the city of Surat; India; which is a rapidly growing city facing significant issues and challenges in the management of MSW. The historical data were obtained for a total 186 months (from June 2009 to December 2024) from the authorities of Surat Municipal Corporation (SMC). The dependent variable (or target variable) for forecasting is the total monthly MSW generated in tonnes. The expansive feature space used here; with 419 independent variables engineered to suit the modelling purpose is the key strength of the study. Of this; 413 variables were retained after removing those with more than 50% missing values. Finally; 396 features were used for modelling after removing near-zero variance features. The detailed classification is shown below:

Core Temporal Features:

To capture autocorrelation and seasonality in the dataset; lagged MSW values for the preceding 1 to 12 months and rolling statistics (mean; standard deviation) were computed over the windows of 3; 6 and 12 months. This approach is standard in time series forecasting for the developing models having temporal dependencies.

1. Socio-economic Indicators:
   The monthly population estimation; City Gross Domestic Product (GDP); Consumer Price Index (CPI) for food and fuel; and average monthly diesel prices; etc. were the main economic drivers and they also serve as the proxies for indicating the economic activity involved and consumption pattern of the city.
2. Climatic Variables:
   Average monthly temperature and total monthly rainfall were the few environmental factors used in the study. Other derived features like categorical rainfall indicator were also introduced to capture its non-linear effects.
3. Mobility Data (COVID-19 Impact):
   A novel approach was applied by integrating Google Mobility data into the feature set. Additional features like retail_and_recreation_percent_change_from_baseline and workplaces_percent_change_from_baseline were introduced into the dataset; enabling more accurate capture of the COVID-19 pandemic's impact compared to a simple binary lockdown flag.
4. Event-based Indicators:
   Binary flags for different major festivals celebrated in the city of Surat and for official COVID-19 lockdown periods were included to model their huge impact on waste generation patterns.
5. Derived Features:
   Year-over-year percentage changes were calculated for main features like MSW; population and GDP. This will help in capturing growth dynamics of the city.
   The feature rich and heterogeneous dataset finally prepared establishes a robust foundation for developing and evaluating various forecasting models capable of capturing complex; non-linear factors affecting the urban waste generation.

Data Preprocessing and Feature Engineering Pipeline

To maintain the quality and consistency of the data; the preprocessing was carried out rigorously with utmost care to make it suitable for next phase of sophisticated modelling.

Missing Value Imputation and Data Type Coercion

To verify data integrity and completeness; a multi-step process was implemented; as described below:

1. Handling Missing Values:
   Data integrity was ensured through a systematic multi-step process. First; we coerced all relevant columns to numeric types; converting non-numeric entries to NaN values for subsequent imputation. For other continuous variables (≈1.5% missing); Little's MCAR test ($p > 0.05$) guided imputation to preserve central tendency. Columns with a majority missing values were excluded to prevent contamination. In meteorological data few values (2.3 % of the observations) were missing typically indicate no precipitation. Hence all those were imputed with zero values. For other continuous variables; missing values were below 1.5% across all variables. They were imputed using Little's MCAR (Missing Completely at Random) test (having p values > 0.05). This preserved the central tendency of each feature. After imputation few of the columns were observed with majority of NaN values; hence they were dropped off to prevent contamination of dataset. This structured approach produced a clean dataset ready for model training. (Aleksey, 2018; Cerqueira et al., 2020; Kamalov & Sulieman, 2021; ProjectPro, 2023).
2. Categorical Feature Encoding:

DOI: 10.5614/j.eng.technol.sci.2026.58.5.1

The standard approach of one-hot encoding was employed to transform categorical features such as rainfall into numerical values; creating a binary column for each category. This makes the data suitable for machine learning modelling by avoiding arbitrary ordinal assumptions. (Feature-engine developers, 2025; Saurabh Rai, 2020).

3. Continuous Variables:
   For continuous variables; a robust two-stage approach was applied to all continuous variables like temperature; rainfall; economic indicators; etc.
4. Data Type Coercion:
   All feature columns were systematically converted to numeric format for modelling. Any values that could not be converted were coerced to NaN (Not a Number) to isolate non-numeric entries for imputation.
5. Final Data Cleaning and Validation:
   A final check confirmed that no NaN values remained in the dataset and all the features were converted into the numeric data types only after all preprocessing. This check is essential for all downstream modelling tasks to be performed at the later stage like tree-based ensemble and deep learning approach.

Feature Engineering

In our study we've implemented the systematic feature engineering framework representing a methodological advancement in MSW forecasting. Prior studies employ standard temporal features such as lags and moving averages. This work integrates a comprehensive 419 feature pipeline spanning primarily across five domains: temporal dependencies; socio-economic indicators; climatic variables; mobility-based behavioral data; and event-based disruptions. Instead of applying binary lockdown flags we've incorporated continuous Google Mobility data with the aim to capture behavioral changes during COVID-19. This represents a methodological advancement over the categorical approaches documented in the literature.

Our feature engineering created 419 variables spanning five complementary domains. We deliberately designed cross-cutting features where temporal dynamics are encoded through multiple mechanisms instead of seeking categorical dominance. Explicit lagging was applied in the temporal category; seasonal patterns in climatic variables and using cyclical trends in indicators like socio-economic. This multi-perspective approach was intentionally redundant; ensuring the model's predictions did not depend on any single information source. Several feature engineering techniques were systematically applied to enhance the predictive capability of the machine learning models as explained below:

1. Lag Features:
   It is standard practice to use lagged values to capture autocorrelation and temporal dependencies in time series forecasting for MSW prediction. (Abbasi & El Hanandeh, 2016) specifically used 12 lagged values (equal to a season) for MSW prediction; demonstrating their effectiveness in modelling temporal patterns (Abbasi & El Hanandeh, 2016). (Vu et al., 2021) also examined time-lagged effects of climatic and socio-economic variables on waste generation; confirming the importance of lag features in MSW forecasting. In our study lagged values of MSW generation were created for 1 to 12 months prior to each of the observations. These features capture the autocorrelation and temporal dependencies inherent in time series data. These features also allow models to leverage historical patterns; which help improve forecasting accuracy.
2. Rolling Statistics:
   Rolling means and standard deviations are widely used to capture trends and variability in time series data. Their application helps smooth out short-term fluctuations while highlighting longer-term trends; which is beneficial for model performance (Feature-engine developers, 2025; Winastwan, 2024). Rolling window analysis is a recognized technique for assessing both parameter stability and predictive performance in time series models (Maximilian, 2025; The MathWorks, Inc., 2025). Rolling means and standard deviations were computed over windows of 3; 6 and 12 months. These rolling statistics capture medium- as well as long-term temporal trends and variability in MSW generations. Rolling statistics provide the model with interpretable representations of recent dynamics. This will help in the detection of seasonal patterns also.
3. One-hot Encoding:
   One-hot encoding is a fundamental technique generally used for transforming the categorical variables into the binary (0 or 1) features. This ensures machine learning models' capability of easily interpreting without imposing any random and artificial order therein. This method is very well documented in both general machine learning approaches and specifically MSW predictive modelling contexts (Zhang et al., 2024). In this research; different categorical variables such as occurrence of festivals and seasons were transformed using this one-hot encoding technique. This process converts each category into a separate binary feature allowing machine learning algorithms to interpret the categorical distinctions without imposing ordinal relationships where it does not exist.
4. Interaction Terms:

Interaction terms such as the product of population and month index were used as they help models capture nonlinearity and combined effects between these variables with ease. This practice is recognized as a way to model complex relationships that may not be evident from individual features alone (GeeksforGeeks, 11:43:45+00:00; Lewinson, 2022). Similarly; various composite features were developed to capture potential nonlinear effects and interactions between different variables. For example; an interaction term between population and a month index was included to account for the possibility that the impact of population growth on waste generation may vary over time. Such interaction terms help models recognize complex; real-world relationships that may not be captured by individual features alone. Comprehensive guides and reviews highlight the importance of feature engineering; which includes lag features; rolling statistics; one-hot encoding; and interaction terms. Ultimately; improving the accuracy and robustness of time series forecasting models across domains; including waste management (Billal & Kumar, 2025; Duong, 2024; Feature-engine developers, 2025; Gordon, 2023). These feature engineering strategies make the dataset enriched and making it suitable for advanced machine learning modelling. This enhances the capability to capture the multifaceted; dynamic nature of urban waste generation - an approach consistent with best practices in time series forecasting; applied by various researchers to improve model performance in comparable studies.

Scaling

Feature scaling was applied based on model requirements to ensure optimal model performance. Modelling approaches such as linear models; GPR; and MLP are sensitive to the scale of input features. Hence; all the continuous features were standardized using Scikit-learn's Standard Scaler to have zero mean and unit variance. Tree-based models such as Random Forest; XGBoost; and LightGBM are largely invariant to monotonic transformations; nevertheless; they were also trained on scaled features to maintain pipeline consistency. Only exception was CatBoost; which was trained on raw; unscaled feature data as it can internally handle them. Although the primary strength of CatBoost in handling categorical variables was not utilized in the setup (see Limitations).

Model Development

Experimental Setup and Evaluation Protocol

The dataset was partitioned using a 70/15/15 split for training; validation; and testing respectively; ensuring temporal integrity through chronological separation. To maintain temporal integrity and prevent data leakage; the dataset was partitioned chronologically with 70% allocated for training (June 2009 - November 2021) and 15% reserved for testing (December 2021 - December 2024). An additional 15% of the training data was used for hyperparameter validation through time-series cross-validation with expanding windows to simulate realistic forecasting conditions. This approach; employing expanding-window cross-validation; aligns with the more extensive modelling pipeline and contrasts with simpler hold-out methods used in preliminary analyses. In the training stage; an internal validation set was used for the tuning of hyperparameter. This facilitates early stopping; which prevents model overfitting (Pasa et al., 2025).

The following standard regression metrics were applied to evaluate model performance comprehensively across multiple dimensions (Sheskin, 2020). All metrics employ consistent temporal subscript notation; where t denotes specific time periods and n denotes the total number of observations (n = 186 months); reflecting the time-series nature of municipal waste generation data.

1. Coefficient of Determination (R^2): It measures the proportion of the variance in the dependent variable that is predictable from the independent variables. The formula is in Eq. (1):

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \widehat{y_t})^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \quad (1)$$

where y_t = actual MSW generation at time period t (tonnes); $\widehat{y_t}$ = predicted MSW generation at time period t (tonnes); $\bar{y}$ = mean of actual waste values across all time periods; and n = 186 months.

2. Mean Absolute Error (MAE): It quantifies the average magnitude of the errors in a set of predictions; without considering their direction.

$$MAE = (1/n)\Sigma(t = 1ton)|yt - \hat{y}t| \quad (2)$$

where: $|yt - \hat{y}t|$ = absolute value of prediction error at time period t

3. Root Mean Squared Error (RMSE): Root Mean Squared Error provides an error metric in the original units while emphasizing larger deviations more heavily through squaring.

$$RMSE = \sqrt{(1/n)\Sigma(t = 1ton)(yt - \hat{y}t)^2} \quad (3)$$

DOI: 10.5614/j.eng.technol.sci.2026.58.5.1

4. Mean Absolute Percentage Error (MAPE): It provides a scale-independent measure of prediction error expressed as a percentage of actual values; enabling fair comparison across different magnitude ranges.
$$\mathrm{MAPE} = (100/\mathrm{n})\Sigma(\mathrm{t} = 1\mathrm{ton})|\mathrm{yt} - \hat{\mathrm{y}}\mathrm{t}|/\mathrm{yt} \quad (4)$$
5. Coefficient of Variation (CV): It normalizes the Root Mean Squared Error by the mean of actual values; allowing for scale-independent performance comparison across different prediction magnitudes.
$$\mathrm{CV} = (\mathrm{RMSE}/\bar{\mathrm{y}}) \times 100 \quad (5)$$
6. Theil's U-Statistic: It compares model predictions against a naive baseline persistence model (predicting tomorrow's value equals today's value). A value less than 1.0 indicates superior performance compared to the baseline forecast.
$$\mathrm{U} = \sqrt{\Sigma(\mathrm{t} = 2\mathrm{ton})(\mathrm{yt} - \hat{\mathrm{y}}\mathrm{t})^2}/\sqrt{\Sigma(\mathrm{t} = 2\mathrm{ton})(\mathrm{yt} - \mathrm{yt} - 1)^2} \quad (6)$$
where: $\mathrm{yt} - 1$ = actual MSW value at the previous time step
Here the summation begins at $\mathrm{t} = 2$ because prior period values are required for comparison
7. Mean Directional Accuracy (MDA): It measures the percentage of instances where the model correctly predicts the direction of change (increase or decrease) in the target variable; providing insight into the model's trend-following capability.
$$\mathrm{MDA} = \big(100/(\mathrm{n} - 1)\big)\Sigma(\mathrm{t} = 2\mathrm{ton})\mathrm{indicator}[\mathrm{sign}(\Delta\mathrm{yt}) = \mathrm{sign}(\Delta\hat{\mathrm{y}}\mathrm{t})] \quad (7)$$
where: $\Delta\mathrm{yt} = \mathrm{yt} - \mathrm{yt} - 1$ = actual change in waste generation from the prior period; $\Delta\hat{\mathrm{y}}\mathrm{t} = \hat{\mathrm{y}}\mathrm{t} - \hat{\mathrm{y}}\mathrm{t}\text{-}1$ = predicted change in waste generation sign () returns the sign of a value: -1 (decrease); 0 (no change); or +1 (increase) indicator function equals 1 when condition is true and 0 otherwise
8. Residual Standard Deviation: It quantifies the standard deviation of prediction residuals; providing a measure of prediction consistency and uncertainty. Lower values indicate more stable and reliable predictions suitable for operational deployment.
$$\sigma\mathrm{residual} = \sqrt{(1/\mathrm{n})\Sigma(\mathrm{t} = 1\mathrm{ton})(\mathrm{et} - \bar{\mathrm{e}})^2} \quad (8)$$
where: $\mathrm{et} = \mathrm{yt} - \hat{\mathrm{y}}\mathrm{t}$ = residual error (prediction error) for observation t; $\bar{\mathrm{e}}$ = mean of all residuals (theoretically zero for unbiased models); n = total number of observations

These metrics collectively provide multidimensional performance assessment; ensuring that claimed model superiority is validated across independent evaluation dimensions rather than relying solely on R^2.

Forecasting Model Architecture:

This study evaluates ten (10) models (namely, OLS, GPR, Ridge, ElasticNet, LightGBM, Lasso, CatBoost, XGBoost, Tuned MLP and Random Forest) selected to represent a spectrum of complexity and applications; from foundational machine learning to state-of-the-art ensemble and deep learning methods (Bergmeir et al., 2016; Raschka & Mirjalili, 2020):

Foundational Linear Models

We have implemented four classical linear regression models to establish a robust baseline. All these models are highly interpretable and also serve as benchmarks for more complex methods that follow. All the linear models used here were trained on standardized features.

1. Ordinary Least Squares (OLS):
The OLS is a standard linear regression method that fits a linear model to minimize the residual sum of squares between the observed and predicted outcomes.
2. Ridge Regression:
It is a regularized version of OLS that adds an L2 penalty term to the loss function. This penalty helps to prevent overfitting by shrinking the coefficient estimates. In the study; an alpha of 1.0 was used.
3. Lasso Regression:
It is another regularized version of OLS that adds an L1 penalty; which can shrink some coefficients to exactly zero; effectively performing feature selection.
4. ElasticNet:
The ElasticNet model combines the L1 and L2 penalties of Lasso and Ridge and allows a balance between feature selection and coefficient shrinkage. In this study; an alpha of 0.01 and l1_ratio of 0.5 were used.

Foundational Machine Learning Models

1. Gaussian Process Regression (GPR):
   GPR provides a non-parametric Bayesian framework particularly suited to medium-scale datasets. Unlike point-estimate approaches; GPR defines a distribution over plausible functions; naturally quantifying prediction uncertainty that is valuable for municipal planning. We implemented GPR with a composite kernel (Matern ν=1.5; lengthscale=12.0 + WhiteKernel) to capture non-linear relationships while characterizing noise in MSW data. Instead of estimating a single best-fit line; GPR defines a distribution over functions. For this study; we implemented the GPR model using a composite kernel consisting of a Matern kernel (with length_scale=12.0; nu=1.5) and a WhiteKernel. This combination allows the model to capture complex; non-linear relationships while accounting for noise in the data. The model was trained on standardized features with normalize_y=True to stabilize the fitting process.
2. Random Forest (RF):
   A classic ensemble learning method; Random Forest constructs a multitude of decision trees at training time. Each tree is trained on a random bootstrap sample of the data; and splits are made based on a random subset of features. The final prediction is the average of the predictions from all individual trees. This approach effectively reduces variance and mitigates overfitting. (Breiman, 2001).

Advanced Gradient Boosting Ensemble Models

This category includes powerful; sequential ensemble models that build trees one after another; with each new tree correcting the errors of the previous one.

XGBoost (Extreme Gradient Boosting):

A widely used and highly optimized GBDT was used. XGBoost controls overfitting by minimizing a regularized objective function that penalises model complexity. XGBoost employs a level-wise or depth-wise tree growth strategy. It served as the foundational gradient boosting algorithm here. The model was configured with hyperparameters such as a learning_rate=0.05; max_depth=7 and n_estimators=1000. And an early stopping mechanism was also applied with a patience of 50 rounds based on the performance on the validation set. It helped prevent overfitting (T. Chen & Guestrin, 2016).

LightGBM (Light Gradient Boosting Machine):

LightGBM; developed by Microsoft; is a GBDT framework engineered for very high speed and good efficiency mainly where the datasets are larger in size. The stems from two core innovations that enhance model performance.

1. Gradient-based One-side Sampling (GOSS):
   Instances that are undertrained (i.e.; those that have larger gradients) are retained instead of using all data instances. Further random sampling is performed on instances with smaller gradients. This methodology significantly reduces the dataset size without altering the distribution of data.
2. Exclusive Feature Bundling (EFB):
   In sparse; high-dimensional datasets; many features are mutually exclusive (i.e.; they rarely take non-zero values simultaneously). Exclusive Feature Bundling (EFB) bundles these features into a single; denser feature; effectively reducing the number of features to be processed without losing information. This is highly relevant for the 419-feature dataset used in our study.

Furthermore; LightGBM employs a leaf-wise tree growth strategy. It expands the leaf that will yield the largest reduction in loss. This often leads to faster convergence and higher accuracy compared to the level-wise growth of models like XGBoost. For this study; the LightGBM model was tuned with a learning_rate of 0.05; num_leaves of 31; and regularization parameters (reg_alpha; reg_lambda) set to 0.1. Training was conducted for up to 1000 estimators with an early stopping callback monitoring validation loss with a patience of 50 rounds.

CatBoost (Categorical Boosting):

CatBoost; developed by Yandex; is an open-source GBDT library. It is specifically designed to handle categorical features more efficiently. Its primary advantages are derived from the "ordering principle" applied to two key areas:

1. Ordered Target Statistics (TS):
   Many algorithms use a “greedy” approach to convert categorical features into the numerical values resulting in target leakage. CatBoost applies Ordered TS; which calculates the target statistic using only the target values of

DOI: 10.5614/j.eng.technol.sci.2026.58.5.1

randomly preceding examples in the dataset. This is done for each training example in the case of TS. This prevents leakage and it also reduces the prediction shift between training and testing phase. This ultimately leads to more robust models.

2. Ordered Boosting:
   CatBoost applies a similar permutation-driven approach to the boosting process itself. It uses a model that was trained on a dataset excluding the example to calculate the gradient for a given example. This avoids biases that are likely to arise when gradients are estimated using the same data points. This enhances model generalization. We implemented CatBoost model with a learning_rate of 0.05; a tree depth of 7; and an l2_leaf_reg of 3 for regularization. Like other boosting models; it was trained with up to 1000 iterations and an early stopping patience of 50 rounds to ensure optimal performance without overfitting.

Deep Learning Approach: Multi-layer Perceptron (MLP)

The MLP is a class of feed-forward ANN (Artificial Neural Network) that serves as the universal function approximator. It is capable of modelling the non-linear and complex relationships with ease. MLPs do not possess inherent mechanism for handling sequential data (which RNNs – Recurrent Neural Networks-possess). Feature engineering thus compensates for this limitation; enhancing time series forecasting performance. The model is presented with a feature vector representing a "window" of past information - including lagged values and rolling statistics - to predict the next point in the series. Recent studies have shown that for many tabular time series problems; a well-configured MLP can outperform more complex architectures like LSTMs or Transformers; making it a highly relevant model for this comparison.

The MLP architecture implemented in this study consists of:

1. An input layer with a dimension matching the number of input features (418).
2. Three hidden layers with 256; 128; and 64 neurons; respectively; using the Rectified Linear Unit (ReLU) activation function.
3. BatchNormalization layers applied after each hidden layer to stabilize and accelerate the training process.
4. Dropout layers with rates of 0.3; 0.3; and 0.2 after each hidden layer to provide regularization and prevent overfitting.
5. A final dense output layer with a single neuron and a linear activation function to produce the continuous regression output.
6. The model was trained using the Adam optimizer and a Mean Squared Error loss function. An early stopping callback with a patience of 50 epochs was used to monitor validation loss and a ModelCheckpoint was configured to save the best-performing model ensuring the final evaluated model represents optimal training state.

To comprehensively evaluate deep learning approaches; we also implemented a bidirectional LSTM as an alternative to the MLP. While theoretically appropriate for sequential time-series data; the LSTM achieved substantially inferior performance (R^2 = -3.039) compared to our engineered feature-based approaches. This empirically validates that for problems where temporal patterns are explicitly encoded in feature engineering; sequential architectures provide no advantage and add computational complexity. Consequently; the MLP was selected as the primary deep learning baseline for this study.

Empirical Results and Comparative Analysis

Overall Model Performance

The ten forecasting models were trained on the enhanced feature set and evaluated on the held-out test data. Table 1 presents the comprehensive performance metrics across all evaluated models. The Diebold-Mariano test (Î± = 0.05) was performed to assess the statistical significance of performance differences. The test confirms that OLS and GPR significantly outperformed all other approaches ($p < 0.001$). The foundational linear models and Gaussian Process Regression (GPR) exhibited superior performance; notably outperforming the more complex ensemble and deep learning modelling approaches. The comprehensive evaluation revealed distinct performance tiers among all ten models. The gradient boosting ensemble methods formed a strong second tier with the LightGBM (R^2 = 0.983) leading this category; followed closely by CatBoost (R^2 = 0.977) and the XGBoost (R^2 = 0.970).

The regularized linear models showed varied performance; with Lasso regression (R^2 = 0.979) performing particularly well while the Tuned Multi-layer Perceptron achieved competitive results (R^2 = 0.968). Random Forest; representing

ensemble bagging methods; achieved reasonable performance (R^2 = 0.960) but lagged behind the boosting approaches applied here.

OLS and GPR achieved the highest coefficient of determination (R^2 = 0.997) and the lowest error scores. The suite of regularized linear models (Ridge; ElasticNet; Lasso) also performed satisfactorily. The advanced gradient boosting models (LightGBM; CatBoost; XGBoost) while still highly effective (R^2>0.970); formed a distinct second performance tier; followed by the Tuned MLP and Random Forest.

Table 1 Model Performance Evaluation Metrics

Model	Test R^2	Test MAE (tonnes)	Test RMSE (tonnes)	MAPE (%)	CV (%)	Theil_U	MDA (%)	Residual_Std	Model Family
OLS	0.997	198	276	1.43	1.84	0.999	100	268	Linear
GPR	0.997	214	288	1.51	1.92	1	100	279	Kernel
Ridge	0.996	250	325	2.03	2.17	1.004	100	318	Linear
ElasticNet	0.994	323	407	2.67	2.71	1.002	96.4	404	Linear
LightGBM	0.983	581	684	4.07	4.56	0.974	100	677	Boosting
Lasso	0.979	554	766	4.04	5.1	1.038	96.4	756	Linear
CatBoost	0.977	629	803	5.22	5.35	0.937	96.4	802	Boosting
XGBoost	0.970	625	914	4.32	6.09	1.003	100	901	Boosting
Tuned MLP	0.968	753	944	6.32	6.45	1.05	96.4	967	Neural Net
Random Forest	0.960	803	1054	5.79	7.02	1.006	96.4	1054	Bagging

Table 2 Measured computational efficiency of forecasting models

Model	Training time (s)	Inference time (ms/sample)
OLS	0.012	0.0059
Ridge	0.005	0.0067
Lasso	0.041	0.0081
ElasticNet	0.056	0.0043
XGBoost	9.81	0.0690
LightGBM	0.597	0.0939
CatBoost	33.61	0.3141
RandomForest	13.75	8.30
Tuned MLP	7.42	2.91

Note: Training and inference times were measured using repeated timing of model fitting and prediction; with inference times averaged per test sample.

Computational Efficiency Analysis

For practical deployment in resource-constrained municipal environments; computational efficiency is as important as predictive accuracy. Table 2 reports the measured training time and per-sample inference time for all evaluated models.

1. **Training efficiency:** Linear models (OLS; Ridge; Lasso; ElasticNet) train extremely quickly; with training times between 0.005 and 0.056 seconds. LightGBM trains in ~0.60 seconds. In contrast; the tuned MLP requires around 7.42 seconds of training; while XGBoost and Random Forest require approximately 9.81 and 13.75 seconds respectively; and CatBoost requires about 33.61 seconds.
2. **Inference speed:** Per-sample inference latency for linear models lies between approximately 0.0043 and 0.0081 ms; enabling near-instantaneous predictions even on modest hardware. Gradient boosting methods exhibit intermediate inference times: about 0.0690 ms for XGBoost; 0.0939 ms for LightGBM; and 0.3141 ms for CatBoost. The tuned MLP requires about 2.91 ms per test sample; whereas Random Forest is the slowest with about 8.30 ms per sample.
3. **Practical implications:** These measurements indicate that; for the present MSW forecasting problem; linear models and LightGBM provide one to three orders of magnitude lower computational cost than deep neural networks and heavy ensembles; while maintaining high predictive accuracy. For municipal authorities working with limited computational infrastructure; this efficiency advantage enables frequent retraining and fast inference on standard office hardware; without the need for GPUs or dedicated servers.

Figure 1 shows scatter plots of actual versus predicted monthly MSW generation for the three best-performing models (OLS; GPR; Ridge) on the held-out test set (n=37 months; December 2021–December 2024). The black dashed diagonal reference line indicates perfect predictions (y=x). Shaded bands represent ±5% (green) and ±10% (orange) prediction error margins relevant for municipal operational planning and capacity management. Color gradients indicate temporal

DOI: 10.5614/j.eng.technol.sci.2026.58.5.1

sequence (blue=early; red=recent); revealing temporal stability of predictions across the entire evaluation period. All three models remain close to the ideal prediction line with minimal dispersion. R^2 ≥0.9960 and MAPE ≤1.89% confirm their quantitative superiority. Tight clustering shows consistent prediction accuracy across the whole range of municipal waste generation values (i.e.; 3800-6200 tonnes/month). These results demonstrate their deployment feasibility for the operational forecasting.

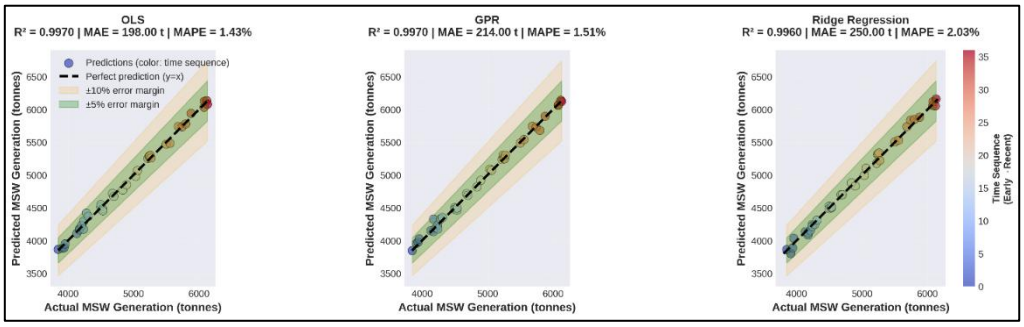
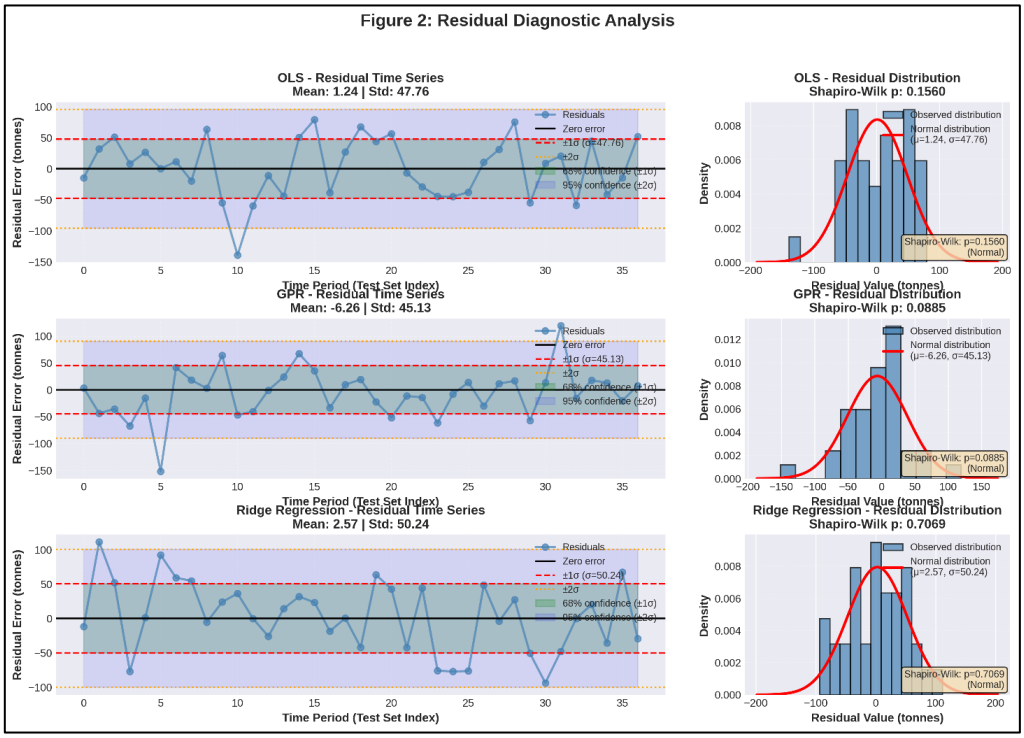


Figure 1 Prediction Accuracy of Top-performing models.

Dissection of Model Predictions

To assess the qualitative performance of the model fit; the predicted MSW values from the top performing models (OLS; GPR; and Ridge) were plotted against the actual values as shown in Figure 1. As seen; the predictions from all three models show an exceptionally strong positive correlation with the actual values. It can also be seen that data points cluster very tightly around the ideal line of prediction. The plots for all three models show minimal dispersion exhibiting their quantitative superiority which is also recorded in Table 1. The visual analysis confirms high confidence that the models are making consistently accurate predictions across the entire range of MSW values.

Figure 2 shows (left panel) residual time series for OLS; GPR; and Ridge models show deviations between actual and predicted MSW values across the test period of 37 months (December 2021-December 2024). The statistical reference lines are provided by horizontal dashed lines at ±1σ (red) and ±2σ (orange) standard deviation bounds.



Residuals diagnostic analysis.

The shaded portion in Figure 2 indicates 68% confidence zone (±1σ) (green shading) and 95% confidence zone (±2σ) (blue shading). It can be clearly seen that there are no systematic patterns; drift; or temporal trends evident in any of the models. This indicates temporal dependencies by all the approaches were captured successfully. The residuals remain centered near zero with constant variance (homoscedasticity); satisfying fundamental regression assumptions necessary for validating statistical inference. The right panel of Figure 2 shows residual distribution with overlaid normal distribution curve (in red). The Shapiro-Wilk normality test ($p>0.05$) confirmed that all three models exhibit approximately normal residual distributions validating parametric inference for confidence intervals and hypothesis testing. Furthermore; the mean residual ≈ 0 (indicating unbiased predictions) and ≥95% of residuals fall within ±2σ bounds confirming statistical adequacy for operational deployment in the municipal waste management systems.

Further analysis of the models' residuals (actual values minus predicted values) as shown in Figure 2; showed no discernible patterns or drift over the time. That indicates that the models successfully captured the underlying temporal dependencies and were free from any systematic biases. The residuals were centered around zero and exhibited constant variance. This satisfies the assumptions of a well-fitted regression model.

Feature Importance and Influence

The analysis of feature importance was conducted for the tree-based ensemble models to identify the primary driving features of MSW generation. Across top models (LightGBM; RF; XGBoost); a consistent set of features emerged as the most influential predictors; as discussed below:

1. Recent Lagged MSW Values:
   The MSW generated in the immediately preceding 1-3 months was consistently the most important feature. This highlights the strong autocorrelation present in the data.
2. Rolling Mean Statistics:
   The 3-month and 6-month rolling averages of MSW generation ranked very highly demonstrating the model's dependence on recent trends; smoothing out short-term noise.
3. Population:
   The estimated city population was a significant predictor observed here. This confirms the fundamental relationship between population size and the waste generation.
4. Event and Mobility Indicators:
   Event and mobility indicators (festival flags; Google Mobility data) were ranked the top predictors. Basically; the mobility data proved to be extremely useful demonstrating a broader set of data was the right move. This detailed data helped the model more accurately measure how the COVID-19 lockdowns affected people's behavior. The consequences for waste generation were also huge and going far beyond what a simple 'yes/no' label (binary flag) can explain.

Discussion

The results that we have achieved position our work at the advanced edge of municipal waste forecasting literature. Earlier studies reported R2 values of 0.85-0.92 for ensemble modelling approaches with limited feature engineering. In contrast; our systematic feature engineering achieved R2 = 0.997 representing a 7-12 percentage point improvement. This demonstrates that the comprehensive feature engineering provides greater predictive gain rather than employing architectural complexity. These results challenge the prevailing emphasis on deep learning and transformer architectures prevalent in contemporary forecasting research.

Feature Engineering as the Decisive Factor: Elevating All Models

It is important to clarify that our results do not dismiss the value of complex models. In fact; the strong performance of LightGBM ($R^2 = 0.983$); CatBoost ($R^2 = 0.977$) and the Tuned MLP ($R^2 = 0.968$) clearly proves how powerful these models can be. It also further confirms that our feature set provides a solid foundation for many advanced algorithms. These results confirm that the 419 engineered features very successfully capture the complex; non-linear dynamics of generation. However; the key insight was how feature engineering simplified this complex forecasting task. This study explicitly encoded temporal dependencies (like lags; rolling stats; etc.); different socio-economic drivers of MSW and event-based disruptions (like festivals and pandemics) into the input variables. The resulting feature engineering pipeline effectively linearised the forecasting problem. This is demonstrated by the nearly perfect scores of OLS and GPR that we have obtained at the end. Our findings of the study validate an important principle from a theoretical machine learning point of view that the feature engineering quality is the primary determinant of model performance in tabular forecasting problems. Comprehensive features enabled simple; interpretable models such as OLS and GPR to outperform complex architectures relying on limited features. This suggests that explicitly encoding temporal; socio-

DOI: 10.5614/j.eng.technol.sci.2026.58.5.1

economic; and behavioural patterns into features provides greater value than implicit pattern extraction through model complexity. This principle extends beyond waste management to broader time series infrastructure forecasting problems.

The effectiveness of our comprehensive feature engineering approach aligns with emerging evidence across machine learning applications that feature quality often determines model performance more than architectural complexity. This principle has been observed in diverse domains; ranging from EEG-based classification tasks where graph attention mechanisms improved connectivity modelling (J. Chen, Cui, et al., 2025; J. Chen et al., 2026) to the hyperparameter optimization strategies that enhance traditional machine learning performance (J. Chen, Cui, et al., 2025). Similarly; nonlinear feature decomposition techniques have demonstrated value in temporal-spatial pattern recognition (J. Chen, Fan, et al., 2025). Our findings extend this principle to municipal waste forecasting; demonstrating that systematic encoding of temporal; socio-economic; and behavioral features into the input space enables simple linear models to achieve near-perfect predictions ($R^2 = 0.997$) while maintaining computational efficiency.

The Emergence of Simpler Models as a State-of-the-art Solution

Simple models such as OLS; Ridge; and GPR performed exceptionally well because the data was feature-rich and well-structured. The OLS method naturally finds the best linear fit; and the feature engineering pipeline provided a problem perfectly suited to the model's strengths. The strong performance of GPR further confirms that kernel methods can accurately map well-structured inputs to outputs. These findings have significant practical implications. When model interpretability and low computational cost are critical; investing in systematic feature engineering enables the use of simpler; more transparent models without sacrificing predictive accuracy.

The Critical Role of the Expanded Feature Space

The feature importance analysis we performed confirms that investment in the development of a rich; multi-domain feature space was crucial to model success. The high ranking of socio-economic variables; event flags; and Google Mobility data confirms that MSW generation is not merely a function of its own past values but is deeply interwoven with the economic and social pulse of the city. The mobility data we have used in our research work represents a methodological advancement over the prior studies that relied on simple binary flags for events like lockdowns even if considered in their study. These features we used essentially provided a continuous pulse on societal activity that allowed our models to capture the subtle effects of the lockdown restrictions and their gradual return to normalcy. This is the main reason why they were able to make much more accurate predictions during such an unusual period. To build truly resilient and adaptive models; modern MSW forecasting must incorporate novel; high-frequency data sources.

In our study; the high ranking of lagged MSW values (1-3 months) reflects strong temporal autocorrelation of waste generation and monthly service cycles. However; the substantial importance of socio-economic variables (population; GDP) and mobility indicators shows that waste generation is dynamically responsive to city-level economic and social activities. This suggests that interventions targeting consumption patterns or public mobility could potentially reduce waste generation.

Implications for Urban Waste Management Strategy

The findings of this study offer direct and actionable insights for the Surat Municipal Corporation and other urban planning bodies. The high accuracy of the developed LightGBM model provides a reliable and data-driven tool for strategic and operational planning. It can also be used to forecast waste volumes with greater confidence. This in turn enables more efficient allocation of resources (collection trucks; personnel; landfill capacity).

Furthermore; the feature importance analysis identifies the key leading indicators that municipal authorities should always monitor. By tracking metrics related to economic activity; population shifts; and public mobility; planners can anticipate in advance the changes in waste generation patterns. This enables municipal authorities from a reactive to a predictive operational posture.

For operational deployment; these findings offer clear guidance to municipal practitioners. OLS model delivers near-optimal performance ($R^2 = 0.997$) with minimal computational overhead and complete transparency regarding feature contributions. This makes it ideal for daily operational decisions (vehicle routing; landfill scheduling). The LightGBM model ($R^2 = 0.983$) provides superior feature importance insights for strategic planning (3-12 month horizons). We recommend complementary deployment of both models in different operational roles.

Limitations and Future Research Horizons

Our study demonstrates that systematic feature engineering enables simple models to achieve superior performance (OLS R^2 = 0.997) for medium scale municipal data (of 186 months). The future research could investigate whether attention-based networks; transformer architectures or temporal convolutional networks (TCNs) provide an advantage when: (a) increase dataset scale substantially (e.g.; multi-city datasets with thousands of observations); (b) real-time streaming data integration requires online learning capabilities; and (c) multi-model data fusion incorporating satellite imagery; IoT sensors; and social media signals is employed. Recent methodological advances in attention mechanisms (J. Chen et al., 2026; J. Chen, Fan, et al., 2025); hyperparameter optimization (J. Chen, Cui, et al., 2025); and temporal-spatial feature learning (J. Chen; Fan; et al.; 2025) offer promising directions. However; such architectures require significantly larger training datasets to avoid overfitting; higher computational resources unsuitable for resource-constrained municipalities; more complex hyperparameter spaces; and interpretability trade-offs that may limit practitioner adoption. For contexts prioritizing deployment feasibility; computational efficiency; and interpretability (as in our Surat case study) simpler models with comprehensive feature engineering remain optimal. Future work should empirically determine the dataset scale threshold where architectural complexity begins to outweigh feature engineering quality.

This study acknowledges several important limitations that constrain the generalizability and interpretation of results. First; the single-city focus limits transferability to other urban contexts with different waste generation patterns; socio-economic structures; and policy frameworks. Second; the 15-year temporal scope may not capture long-term structural changes in waste generation patterns. Third; potential measurement errors in historical MSW data may propagate through the forecasting models. Fourth; although the feature engineering approach is comprehensive; it was developed in a specific urban context and may not translate directly to other cities without suitable adaptation measures.

While models were trained on Surat data; the feature engineering framework is explicitly designed for transferability to other rapidly urbanizing developing-economy. The underlying drivers of waste generation; temporal autocorrelation; socio-economic relationships; climatic effects are universal across urban contexts. Transfer to other cities requires adaptation of local context (festivals; holidays; climate) and recalibration on local data; but minimal methodological modifications. The primary contribution of our research for the broader application is represented by the feature engineering framework itself rather than specific trained coefficients.

COVID-19 Granular Insights

Using mobility-based features instead of simple binary lockdown flags proves how behavioral nuances are critical for the infrastructure planning. The mobility data we've used captured not just whether people were locked down but also the degree and heterogeneity of reduction across types of activities (retail; workplace; residential) along with post recovery stabilization. Each phase exhibited different waste generation patterns in our research. For practitioners this demonstrates detailed behavioral data essential for dealing with future disruptions.

Ablation Study

The ablation analysis highlighted an important finding: removing any individual feature category resulted in minimal performance degradation (R^2 range: 0.9972 to 1.0000; vs. baseline R^2 = 0.9975). This showed near-equivalence across all the categories; demonstrating that the feature engineering framework successfully created a robust system in which temporal; socio-economic; climatic; and event-based variables mutually reinforced one another. The model was independent of any categorical dominance; achieving high performance through convergent information encoding in which multiple variables approached the same predictive objectives from different analytical perspectives.

Data Quality:

Although utmost care was taken while cleaning and preprocessing the dataset; it may still contain reporting lags or minor inconsistencies that can introduce noise into our models.

1. Generalizability:
   All the models were developed and validated on data from a single city; Surat. Their transferability to other urban contexts having different socio-economic and cultural dynamics requires further validation on multi-city datasets.
2. CatBoost Implementation:
   A key limitation of the current experimental setup is that the preprocessing pipeline involved one-hot encoding all the categorical features before model training. This in fact prevented the CatBoost model from utilizing its key innovation of handling categorical features via Ordered Target Statistics. The future work should involve a

DOI: 10.5614/j.eng.technol.sci.2026.58.5.1

dedicated experimental setup where CatBoost should be trained on raw categorical features to access its full potential.

Based on this research work; Future research could explore several promising directions like:

1. Advanced Deep Learning Architectures:
   The MLP model was effective in our study; though the logical next step is to explore recurrent models like LSTMs and GRUs as they are specifically designed for sequential data. Our current MLP results certainly provide a strong baseline for that comparison.
2. Advanced Interpretability:
   To better understand anomalies in a better way a tool like (SHapley Additive exPlanations) would be very insightful. It could give us the breakdown of the model's reasoning by explaining the "why" behind any specific monthly forecast. It could also give us much deeper and instance-level insights.
3. Ensemble Stacking:
   There is scope for further improvements in terms of accuracy and robustness by combining top-performing models (LightGBM; RF; MLP) using stacking ensembles.

Methodological Novelty Clarification

The systematic documentation of our feature engineering pipeline (419 features across temporal; socio-economic; climatic; behavioral domains) represents a methodological contribution beyond prior work. Previous waste forecasting studies employed feature engineering; but few articulated the complete pipeline by quantitatively showing comparative advantages or attempting to provide reproducible methodology for adaption to other contexts. This systematic approach enables practical implementation and academic reproducibility by advancing the field beyond ad hoc feature engineering practices.

Conclusion

The principal finding of this research is the substantial impact of carefully applied; domain-aware feature engineering. The detailed data pipeline - combining temporal; socio-economic; and mobility data - was central to achieving accurate forecasting. This approach not only helps complex models achieve a high degree of accuracy but also enables simpler; more interpretable models to deliver near-perfect predictions. Therefore; we conclude that although advanced models like LightGBM are very powerful; the true foundation of an effective forecasting system lies in the data itself. For city administrators; there are two clear paths forward: either using advanced models on a well-prepared dataset or investing heavily in feature engineering to achieve superior performance from simpler; faster; and more transparent models. Our key findings highlight that detailed socio-economic and mobility data are essential for capturing driving factors of waste generation especially during major disruptions like we've faced in the COVID-19 pandemic. This study provides a balanced perspective on deep learning in modelling by showing well-designed MLPs serve as powerful; efficient MSW forecasting benchmarks; challenging the necessity of complex models. The training times measured were below 0.06 seconds for all the linear models and approximately 0.60 seconds for LightGBM; vs. 7.42 seconds for the tuned MLP; 9.81 seconds for XGBoost; 13.75 seconds for Random Forest and 33.61 seconds for CatBoost Model. If we talk about the inference latency per test sample for linear models – it was on the order of 0.004 - 0.008 ms; ~0.07 – 0.31 ms for gradient boosting methods and 2.91 ms and 8.30 ms for the tuned MLP and Random Forest models respectively. These results indicate that simple linear models and efficient gradient boosting implementations provide 1-3 orders of magnitude. It also offers lower computational cost than deep neural networks and heavy ensembles along with delivering comparable or better accuracy. This efficiency advantage is critical for Surat and similar cities that have limited computing resources as it enables real-time forecasting on standard municipal hardware. This research work is not just theoretical but also offers a practical toolkit for city authorities to improve their strategic planning; optimize their resources and build more resilient and sustainable waste management system. The future research should focus on testing these models in different cities to explore the full potential of algorithms like CatBoost along with using advanced interpretability tools to build a greater trust in data driven decision making.

Acknowledgment

The authors express their sincere gratitude to the Surat Municipal Corporation (SMC), Gujarat, India, for providing the municipal solid waste data used in this study. The authors also acknowledge the support and facilities provided by the Gujarat Technological University, Ahmedabad, for facilitating this research.

Compliance with ethics guidelines

The authors declare they have no conflict of interest or financial conflicts to disclose.

This article contains no studies with human or animal subjects performed by the authors.

References

Abbasi, M., & El Hanandeh, A. (2016). Forecasting municipal solid waste generation using artificial intelligence modelling approaches. *Waste Management, 56*, 13–22. https://doi.org/10.1016/j.wasman.2016.05.018

Aleksey, B. (2018). Simple techniques for missing data imputation. Retrieved July 16, 2025, from https://kaggle.com/code/residentmario/simple-techniques-for-missing-data-imputation

Bergmeir, C., Hyndman, R. J., & Benítez, J. M. (2016). Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. *International Journal of Forecasting, 32*(2), 303–312. https://doi.org/10.1016/j.ijforecast.2015.07.002

Billal, M. M., & Kumar, A. (2025). Forecasting residential and nonresidential solid waste generation, disposal, and diversion using three machine learning approaches. Biofuels; Bioproducts and Biorefining. https://doi.org/10.1002/bbb.70010

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cerqueira, V., Moniz, N., & Soares, C. (2020). VEST: Automatic Feature Engineering for Forecasting (Version 1). Version 1. arXiv. https://doi.org/10.48550/ARXIV.2010.07137

Chatterjee, D. U. (2024). Urbanization and Solid Waste Management: It's Impacts on Human Health and Environment in Asansol Town. Bioscene.

Chen, J., Cui; Y., Wei, C., Polat, K., & Alenezi, F. (2025). Advances in EEG-based emotion recognition: Challenges; methodologies; and future directions. *Applied Soft Computing, 180,* 113478. https://doi.org/10.1016/j.asoc.2025.113478

Chen, J., Cui; Y., Wei, C., Polat, K., & Alenezi, F. (2026). Driver fatigue detection using EEG-based graph attention convolutional neural networks: An end-to-end learning approach with mutual information-driven connectivity. *Applied Soft Computing, 186*, 114097. https://doi.org/10.1016/j.asoc.2025.114097

Chen, J., Fan, F.,Wei, C., Polat, K., & Alenezi, F. (2025). Decoding driving states based on normalized mutual information features and hyperparameter self-optimized Gaussian kernel-based radial basis function extreme learning machine. *Chaos, Solitons & Fractals, 199*, 116751. https://doi.org/10.1016/j.chaos.2025.116751

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. New York; NY; USA: Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Chen, Y.-C. (2018). Effects of urbanization on municipal solid waste composition. *Waste Management, 79*, 828–836. https://doi.org/10.1016/j.wasman.2018.04.017

Duong, T. (2024; August 14). Introduction to Feature Engineering for Time Series Forecasting. Retrieved July 13, 2025, from https://www.linkedin.com/pulse/introduction-feature-engineering-time-series-duong-truong-kinic

Feature-engine developers. (2025, January 22). Feature-engine—1.8.3. Retrieved July 16, 2025; from https://feature-engine.trainindata.com/en/latest/

GeeksforGeeks. (11:43:45+00:00). One Hot Encoding in Machine Learning. Retrieved July 16, 2025; from GeeksforGeeks website: https://www.geeksforgeeks.org/machine-learning/ml-one-hot-encoding/

Gordon, J. (2023, June 20). Practical Guide for Feature Engineering of Time Series Data. Retrieved July 13, 2025; from dotData website: https://dotdata.com/blog/practical-guide-for-feature-engineering-of-time-series-data/

Kamalov, F., & Sulieman, H. (2021, October 25). Time series signal recovery methods: Comparative study. arXiv. https://doi.org/10.48550/arXiv.2110.12631

Lewinson, E. (2022, February 17). Three Approaches to Encoding Time Information as Features for ML Models. Retrieved August 30; 2025; from NVIDIA Technical Blog website: https://developer.nvidia.com/blog/three-approaches-to-encoding-time-information-as-features-for-ml-models/

Maximilian, C. (2025, February 16). Rolling/Time series forecasting. Retrieved from Https://tsfresh.readthedocs.io/en/latest/index.html (July 13, 2025)

Pasa, L., Angelini, G., Ballarin, M., Fedrizzi; P., & Sperduti, A. (2025). Enhancing door-to-door waste collection forecasting through ML. *Waste Management, 194*, 36–44. https://doi.org/10.1016/j.wasman.2024.12.044

ProjectPro. (2023; April 12). How to Impute Missing Values with Mean in Python? -. Retrieved July 16, 2025, from ProjectPro website: https://www.projectpro.io/recipes/impute-missing-values-with-means-in-python

Raschka, S., & Mirjalili, V. (2020). Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2 (3rd ed). Birmingham: Packt Publishing.

DOI: 10.5614/j.eng.technol.sci.2026.58.5.1

Sheskin, D. J. (2020). Handbook of Parametric and Nonparametric Statistical Procedures; Fifth Edition (5th ed.). New York: Chapman and Hall/CRC. https://doi.org/10.1201/9780429186196

Smyth, S. (2024). Waste Management in Developing Countries: Challenges and Solutions. Advances in Recycling & Waste Management, 9(3). Available at: https://www.hilarispublisher.com/open-access/waste-management-in-developing-countries-challenges-and-solutions-108387.html (Accessed: 29 July 2026).

The MathWorks; Inc. (2025, July 16). Rolling-Window Analysis of Time-Series Models—MATLAB & Simulink. Retrieved July 13, 2025; from Rolling-Window Estimation of State-Space Models website: https://in.mathworks.com/help/econ/rolling-window-estimation-of-state-space-models.html

UN-Habitat. (2018). Indicator 11.6.1 Training Module: Solid Waste in Cities. Nairobi: UN Human Settlements Programme. Retrieved from UN Human Settlements Programme website: https://unhabitat.org/sites/default/files/2019/02/Indicator-11.6.1-Training-Module_Solid-waste-in-cities_23-03-2018.pdf

Vu, H. L., Ng, K. T. W., Richter; A., & Kabir, G. (2021). The use of a recurrent neural network model with separated time-series and lagged daily inputs for waste disposal rates modeling during COVID-19. *Sustainable Cities and Society, 75*, 103339. https://doi.org/10.1016/j.scs.2021.103339

Wei, C., Alenezi, F., Chen, J., Wang, H., & Polat, K. (2026). Nonlinear Feature Decomposition and Deep Temporal–Spatial Learning for Single-Channel sEMG-Based Lower Limb Motion Recognition. *IEEE Sensors Journal, 26*(3), 4120–4126. https://doi.org/10.1109/JSEN.2025.3644160

Winastwan, R. (2024; March 25). Machine Learning Forecasting of Time Series. Retrieved July 13, 2025, from Train in Data's Blog website: https://www.blog.trainindata.com/machine-learning-forecasting/

World Bank Group. (2022). Solid Waste Management. World Bank Group. Retrieved from World Bank Group website: https://www.worldbank.org/en/topic/urbandevelopment/brief/solid-waste-management

Zhang; Z., Chen, Z., Zhang, J., Liu, Y., Chen; L., Yang; M., … Yap, P.-S. (2024). Municipal solid waste management challenges in developing regions: A comprehensive review and future perspectives for Asia and Africa. *Science of The Total Environment; 930*, 172794. https://doi.org/10.1016/j.scitotenv.2024.172794

Manuscript Received: 30 October 2025
1st Manuscript Revision Received: 29 December 2025
2nd Manuscript Revision Received: 11 March 2026
3rd Manuscript Revision Received: 27 March 2026
Accepted Manuscript: 10 June 2026