

SMART WASTE MANAGEMENT SYSTEM

INDIVIDUAL THESIS

Jayawardhane K G M S

(IT22213358)

Individual component: ML-Based Municipal Waste Prediction,
Composition Estimation and Circular Economy Revenue Hub

Group ID: R26-IT-126

Report submitted in partial fulfilment of the requirements for the

B.Sc. (Hons) Degree in Information Technology

Specialized in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

October 2026

DECLARATION AND SUPERVISOR CERTIFICATION

Candidate declaration. I declare that this thesis is my own work and does not incorporate, without acknowledgement, material previously submitted for a degree or diploma at any university or institute of higher learning. To the best of my knowledge, it contains no material previously published or written by another person except where acknowledgement is made in the text. I grant the Sri Lanka Institute of Information Technology the non-exclusive right to reproduce and distribute this thesis, in whole or in part, in print, electronic, or other media, while retaining the right to use the content in future work.

Academic integrity and AI disclosure. I have disclosed all use of generative AI and other automated tools in accordance with the module policy. I remain responsible for the accuracy, originality, citations, code, data, analyses, and conclusions in this thesis.

Student name: Jayawardhane K G M S

Student ID: IT22213358

Signature:  Date: 09/10/2026

Supervisor certification. The above candidate carried out this research for the bachelor's degree thesis under my supervision.

Role and name	Signature	Date
Supervisor: Uditha Dharmakeerthi		10/10/26
Co-supervisor: Kanishka Yapa		

ABSTRACT

Municipal solid waste management requires reliable information on how much waste is likely to be generated before collection and processing activities are scheduled. Conventional fixed collection plans provide limited support for responding to weekly variations caused by rainfall, calendar events, population characteristics, and differences between municipal zones. This research developed a machine-learning-based decision-support component for the Kalutara Municipal Council to forecast wet and dry municipal waste, estimate waste composition, support compactor-truck planning, and evaluate potential resource-recovery revenue.

Two independent Extreme Gradient Boosting regression pipelines were developed for wet and dry waste. Historical municipal records from 2023–2025 were combined with rainfall, temporal, zonal, demographic, calendar-event, and previous-week waste features. A four-tier weather retrieval mechanism was also implemented to obtain historical or forecast precipitation automatically. Model evaluation used a chronological 80:20 train-test strategy to reduce temporal look-ahead leakage.

On the reported 126-record test set, the dry-waste model achieved an R^2 of 0.9492, MAE of 1.521 tonnes, RMSE of 2.273 tonnes, and MAPE of 12.33%, corresponding to an 87.67% MAPE-derived predictive accuracy. The wet-waste model achieved an R^2 of 0.8035, MAE of 3.275 tonnes, RMSE of 4.826 tonnes, and MAPE of 19.15%. Predictions were integrated with a 5-tonne compactor planning mechanism, composition estimates, resource-recovery calculations, recommendations, prediction history, and an interactive web dashboard. The findings demonstrate the potential of localized machine-learning forecasts to support more informed municipal waste planning while highlighting the higher uncertainty associated with wet-waste prediction.

Keywords: Municipal solid waste, XGBoost, waste forecasting, circular economy, decision support

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my research supervisor, and co-supervisor, for their guidance, constructive feedback, and continuous support throughout this research.

I am also grateful to the academic staff of the Department of Information Technology, Sri Lanka Institute of Information Technology, for the knowledge and resources provided during the research period. My appreciation is extended to the Kalutara Municipal Council and the individuals who supported access to the domain knowledge and municipal waste information required for this study.

I would also like to thank the members of Research Group R26-IT-126 for their collaboration in developing the overall Smart Waste Management System and for coordinating the interfaces between the individual research components.

Finally, I am grateful to my family, friends, and everyone who supported me during the completion of this research.

TABLE OF CONTENTS

1. INTRODUCTION	1
1.1 Overall Project Context.....	1
1.2 Individual Component Context.....	1
1.3 Problem Statement	2
1.4 Research Gap	2
1.5 Aim and Objectives.....	3
1.6 Research Questions or Hypotheses	4
1.7 Scope, Limitations and Assumptions.....	4
1.8 Individual Contribution.....	4
1.9 Thesis Structure.....	5
2. LITERATURE REVIEW AND RELATED WORK	6
2.1 Review Method and Scope.....	6
2.2 Domain and Theoretical Foundations	6
2.2.1 Statistical forecasting methods.....	6
2.2.2 Machine learning methods	6
2.2.3 Factors influencing waste generation.....	7
2.3 Related Research.....	7
2.4 Existing Systems, Products and Techniques.....	8
2.5 Gap Synthesis and Conceptual Framework	8
3. RESEARCH METHODOLOGY	9
3.1 Research Design.....	9
3.2 Data, Participants or Study Materials.....	9
3.3 Experimental or Study Procedure	10
3.3.1 Automated rainfall retrieval	10

3.3.2 Preprocessing and pipeline.....	10
3.3.3 Model formulation	10
3.3.4 Validation strategy	11
3.4 Measures and Evaluation Metrics	11
3.5 Data Analysis	12
3.6 Ethics, Privacy, Security and Responsible AI.....	12
3.7 Reproducibility and Research Artefacts.....	12
4. PRODUCT AND SOLUTION DESIGN.....	13
4.1 Requirements.....	13
4.2 Component Architecture within the Complete Solution.....	13
4.3 Design Alternatives and Decisions	14
4.4 Detailed Design.....	15
4.4.1 Fleet requirement planning	15
4.4.2 Waste composition estimation	15
4.4.3 Circular economy revenue model	16
4.4.4 Recommendation rules.....	16
4.5 Standards and Professional Practice.....	17
5. IMPLEMENTATION, INTEGRATION AND TESTING.....	18
5.1 Development Environment and Technology Stack	18
5.2 Component Implementation.....	18
5.2.1 Back-end services.....	18
5.2.2 Front-end dashboard.....	19
5.3 Integration with the Other Components.....	19
5.4 Verification and Software Testing	19
5.5 Deployment and Operation	20
6. RESULTS AND VALIDATION.....	21

6.1 Research Results	21
6.2 Product and Technical Results	22
6.3 User or Stakeholder Validation.....	22
6.4 Integration Results	22
6.5 Summary of Findings.....	23
7. DISCUSSION	24
7.1 Interpretation of Findings.....	24
7.2 Comparison with Literature Baselines and Existing Solutions.....	24
7.3 Research Contribution.....	25
7.4 Product Contribution.....	25
7.5 Limitations and Threats to Validity	25
7.6 Implications.....	26
8. COMMERCIALISATION AND SUSTAINABILITY	27
8.1 Value Proposition and Target Users	27
8.2 Market and Competitive Position.....	27
8.3 Business or Sustainability Model.....	27
8.4 Scalability, Intellectual Property and Compliance.....	28
8.5 Contribution to the Sustainable Development Goals	28
9. CONCLUSION AND FUTURE DIRECTIONS	29
9.1 Conclusion	29
9.2 Achievement of Objectives	29
9.3 Individual Research and Product Contributions	29
9.4 Future Directions.....	30
REFERENCES.....	31
GLOSSARY.....	33
APPENDICES	34

Appendix A: Extended Results and Supplementary Specifications	34
Appendix B: Academic Integrity and AI Use Disclosure.....	34
Appendix C: System Implementation Evidence	35
C.1 Dataset Evidence	35
C.2 Machine Learning Model Evidence	36
C.3 Dashboard Evidence	37

LIST OF FIGURES

	Page
Figure 4.1 Architecture of the waste prediction and resource monetization component.....	14
Figure 6.1 Predictive accuracy by zone and waste stream on the test set.....	22
Figure C.1: Sample records from the municipal waste dataset used for model development.....	35
Figure C.2: Implementation of the machine learning prediction pipeline	36
Figure C.3: Main waste prediction dashboard	37

LIST OF TABLES

	Page
Table 2.1 Forecasting approaches considered for weekly waste prediction	7
Table 3.1 Collection zones in the study area	9
Table 3.2 Model configuration.....	11
Table 4.1 Functional and non-functional requirements	13
Table 4.2 Main design decisions.....	14
Table 4.3 Assumed waste composition by area type (%)	15
Table 4.4 Resource recovery routes and benchmark rates	16
Table 5.1 Technology stack	18
Table 5.2 Main REST endpoints.....	18
Table 5.3 Verification measures	19
Table 6.1 Overall model performance on the test set (n = 126).....	21
Table 6.2 Zone-level performance on the test set	21
Table 8.1 Illustrative revenue per tonne of dry waste	27
Table 9.1 Achievement of objectives.....	29
Table A.1 Dataset schema.....	34
Table A.2 Climatological mean weekly rainfall used as fallback (mm).....	34

LIST OF ABBREVIATIONS

Abbreviation	Meaning
API	Application Programming Interface
ARIMA	Auto-Regressive Integrated Moving Average
CEA	Central Environmental Authority
ERA5	Fifth-generation ECMWF atmospheric reanalysis
HDPE	High-density polyethylene
IoT	Internet of Things
KMC	Kalutara Municipal Council
LKR	Sri Lankan rupee
MAE	Mean absolute error
MAPE	Mean absolute percentage error
ML	Machine learning
MSW	Municipal solid waste
PET	Polyethylene terephthalate
PP	Polypropylene
R ²	Coefficient of determination
RDF	Refuse-derived fuel
REST	Representational State Transfer
RMSE	Root mean squared error
SDG	Sustainable Development Goal
XGBoost	Extreme Gradient Boosting

1. INTRODUCTION

1.1 Overall Project Context

Municipal solid waste (MSW) management is one of the largest and most visible services that a local authority provides. Global waste generation was estimated at 2.01 billion tonnes in 2016 and is projected to reach 3.40 billion tonnes by 2050 if current trends continue [1]. In Sri Lanka, collection and transport are the responsibility of local authorities such as municipal councils, urban councils and Pradeshiya Sabhas, while the Central Environmental Authority (CEA) leads national policy and programmes on waste management [2]. Local authorities are reported to spend a large share of their operating budgets on collection and transport, yet collection remains largely reactive. Compactor trucks follow fixed routes on fixed days, regardless of how much waste has accumulated in each area.

The overall research project, titled Smart Waste Management System, investigates how digital technologies can improve this process. Instead of treating waste collection as a fixed and purely reactive activity, the project is intended to provide data-supported information for monitoring, prediction, collection planning, and operational decision-making.

Within the overall group solution, different research components address complementary aspects of smart waste management. This thesis concentrates on the predictive analytics component. It produces forward-looking estimates of municipal waste generation that can subsequently support operational actions such as fleet scheduling, resource recovery, and potentially integration with smart-bin monitoring and route-optimization components.

1.2 Individual Component Context

The individual research component provides the forecasting and decision-support layer of the system. It predicts the wet (biodegradable) and dry (recyclable and residual) waste that each of four KMC collection zones will generate in a coming week, using historical collection records, the previous week's tonnage, calendar information

and rainfall. The predicted tonnages are then converted into three operational outputs: the number of 5-tonne compactor truck trips required, an estimate of the material composition of the waste, and the revenue that could be earned if the recoverable fractions are sold as compost or recyclables. The outputs are presented to municipal health inspectors and logistics supervisors through a web dashboard.

Within the group project, the component supplies the zone-level, weekly view of expected demand on the collection service. This view complements the real-time fill-level information from smart bins and the tonnage inputs that a route optimizer needs.

1.3 Problem Statement

Collection planning in the study area is currently based on experience and fixed schedules. Three connected problems follow from this practice.

- No quantitative foresight. Supervisors have no localized model of how much waste each zone will generate in the coming week, so they cannot plan crews and vehicles in advance.
- Poor fleet utilization. Fixed dispatch means that a 5-tonne compactor truck may travel under-loaded, wasting fuel and operator hours, or be overloaded, which increases mechanical wear and the risk of spillage.
- Unrealized resource value. Organic and recyclable fractions are largely sent to disposal without segregation planning, so the council forgoes potential revenue from composting and material recovery.

The research problem is therefore how to forecast zone-level weekly waste generation accurately enough, using data that a council can realistically obtain, so that fleet dispatch and resource-recovery decisions can be made before collection begins.

1.4 Research Gap

Machine learning has been applied to MSW forecasting in several regions [3], but three gaps remain for a Sri Lankan municipal council. First, much of the published work targets cities with different waste streams, climates and cultural calendars, and relatively little reports results from a coastal South Asian local authority using real collection records. Second, forecasting studies commonly stop at reporting forecast

accuracy; they do not connect the forecast to fleet capacity planning or to the economic value of the recovered materials, which are the decisions a council actually needs to make. Third, rainfall and holiday effects are important in the tropics, but forecasting tools often rely on manually entered weather data, which is a barrier to routine use by council staff.

This research addresses these gaps by building and evaluating a localized forecasting framework for KMC that retrieves weather data automatically and passes its predictions to truck-trip, composition and revenue calculations.

1.5 Aim and Objectives

Aim

To design, implement and evaluate a machine learning-based platform that forecasts weekly wet and dry waste generation for the zones of the Kalutara Municipal Council and converts the forecasts into fleet requirements and circular-economy revenue estimates.

The specific objectives are:

1. To train separate gradient-boosted regression models for wet and dry waste using historical zonal waste and environmental data.
2. To automate the retrieval of rainfall inputs so that live forecasts or archived records are supplied to the models without manual meteorological entry.
3. To estimate the organic, plastic, paper and metal composition of the waste from the characteristics of each zone.
4. To calculate the weekly truck trips and vehicle utilization required, given a 5-tonne compactor capacity and the council fleet size.
5. To estimate potential revenue from compost and recyclable materials using localized benchmark prices.
6. To provide an interactive dashboard that gives supervisors predictions, retrospective back testing and rule-based alerts.

1.6 Research Questions or Hypotheses

The study is guided by three research questions.

- RQ1. How accurately can weekly wet and dry waste generation be forecast at zone level from lagged tonnage, rainfall, calendar and zone characteristics?
- RQ2. How does forecast accuracy differ between the wet and dry streams and between zones, and what might explain the differences?
- RQ3. Are the forecast errors small enough, relative to a 5-tonne truck, to support practical dispatch planning?

In relation to RQ2, the working hypothesis (H1) is that dry waste is more predictable than wet waste, because it is tied to relatively stable commercial and packaging patterns whereas wet waste is also affected by moisture and household practices.

1.7 Scope, Limitations and Assumptions

The study is limited to the four collection zones of KMC (Kalutara North, Kalutara South, Katukurunda 1 and Katukurunda 2) and to a weekly forecasting cycle that matches crew and fleet rostering. Waste is divided into two collection streams: wet waste (kitchen and garden waste) and dry waste (plastics, paper and cardboard, glass, metals and non-recyclable residue). The following assumptions apply:

- a standard compactor truck carries 5.0 tonnes and the council operates a fleet of three vehicles;
- material composition is determined by the area type of a zone and is not measured weekly;
- recovery rates and market prices are fixed benchmark values rather than live market data;
- the previous week's collected tonnage is available when a forecast is requested.

1.8 Individual Contribution

My individual contribution to the Smart Waste Management System was the design and implementation of the municipal waste prediction, composition estimation, fleet-planning, and resource-monetization component. It comprises the following work:

- preparation of the zonal weekly dataset and design of the feature set;
- implementation and evaluation of the two XGBoost pipelines (wet and dry waste);
- implementation of the four-tier rainfall retrieval mechanism;
- design of the fleet-planning, composition and revenue calculations;
- development of the Flask back-end and React dashboard, including back testing and recommendation features.

1.9 Thesis Structure

Chapter 2 reviews the literature on waste forecasting and resource recovery. Chapter 3 describes the research methodology, including the data, rainfall retrieval, models and evaluation measures. Chapter 4 presents the requirements and the design of the solution. Chapter 5 explains the implementation, integration and verification. Chapter 6 reports the results, and Chapter 7 discusses them, including limitations. Chapter 8 considers commercialization and sustainability, and Chapter 9 concludes the thesis and outlines future work.

2. LITERATURE REVIEW AND RELATED WORK

2.1 Review Method and Scope

The review focuses on four areas relevant to the research: municipal waste-generation forecasting, machine-learning methods for tabular forecasting problems, explanatory variables affecting municipal waste generation, and data-driven decision support for waste recovery and circular-economy planning.

Previous reviews have shown that municipal waste-generation models differ considerably according to geographical scale, selected waste stream, independent variables, and modelling technique. For this reason, the literature considered for the present study was interpreted with particular attention to whether the proposed approaches could support zone-level and short-term municipal decision-making.

2.2 Domain and Theoretical Foundations

2.2.1 Statistical forecasting methods

Early work on waste generation relied on time-series and regression techniques such as Auto-Regressive Integrated Moving Average (ARIMA) models, multiple linear regression and exponential smoothing [4]. These methods are light to compute and easy to interpret, but they assume linear relationships. They therefore have difficulty representing abrupt holiday effects and interactions between weather and the characteristics of an area.

2.2.2 Machine learning methods

Non-parametric machine learning models can capture non-linear relationships without a specified functional form. Artificial neural networks are flexible but can over fit when the dataset is small and are difficult to interpret. Random forests average many decision trees trained on bootstrap samples [5], which makes them robust to outliers, although tree ensembles cannot extrapolate beyond the range of the training targets. Gradient boosting builds trees sequentially so that each new tree reduces the residual error of the ensemble [6]. XGBoost (Extreme Gradient Boosting) implements this idea with a second-order approximation of the loss and explicit regularization of tree complexity, and it handles sparse and tabular inputs efficiently [7]. A comparison on

medium-sized tabular datasets found that tree-based models remained competitive with, and often better than, deep learning models [8]. Table 2.1 summarizes these approaches.

Table 2.1: Forecasting approaches considered for weekly waste prediction

Approach	Strength	Limitation	Fit for this study
ARIMA / regression	Simple, interpretable	Linear; weak with holiday and weather interactions	Low
Artificial neural network	Models non-linear effects	Over fits small data; opaque	Medium
Random forest	Robust to outliers	Cannot extrapolate beyond training range	Medium
Gradient boosting (XGBoost)	Regularized; strong on tabular data	Needs lagged inputs; limited extrapolation	High

Source: synthesis of [4]–[8]

2.2.3 Factors influencing waste generation

Three groups of factors are usually considered. Meteorological factors such as rainfall change the moisture content and weight of organic waste and can alter household disposal habits; in coastal Sri Lanka the south-west and north-east monsoons produce pronounced seasonal differences. Temporal factors such as public holidays, full-moon Poya days and religious festivals create short-term surges, particularly in commercial areas. Spatial and demographic factors also matter: commercial town centers generate more paper and packaging, beach areas generate tourism-related plastic bottles, and dense residential areas generate large volumes of food waste.

2.3 Related Research

Abbasi and El Hanandeh [3] evaluated several artificial intelligence models for forecasting municipal solid waste generation, illustrating that data-driven models can be applied to this problem with socio-economic and temporal inputs. Their emphasis, like that of much of the literature, is on forecast accuracy for a single city. The choice of tree-based learning for tabular inputs in the present study is consistent with the findings reported in [8].

2.4 Existing Systems, Products and Techniques

In Sri Lanka the CEA promotes aerobic windrow composting and material recovery facilities through national programmes such as Pilisaru and Mihisaru [2]. These facilities depend on a predictable supply of segregated material, but councils usually have no advance estimate of the weekly inbound quantity. Planning of collection is typically based on fixed routes and schedules. Commercial smart-bin platforms generally rely on fill-level sensors and report the current state of containers; they do not forecast zone-level tonnage over a planning horizon or link it to local market prices. The component reported here is therefore intended to complement such sensing rather than replace it.

2.5 Gap Synthesis and Conceptual Framework

The review shows that machine learning is an appropriate technique for waste forecasting, that rainfall, calendar effects and zone characteristics are relevant inputs, and that Sri Lankan resource-recovery programmes would benefit from predictable tonnage. What has not been shown is an integrated, locally validated workflow that moves from forecasts to dispatch and revenue decisions with minimal manual input. The conceptual framework of this research therefore has four stages: (1) acquire historical records and automatically retrieved rainfall; (2) forecast wet and dry tonnage per zone; (3) translate tonnage into truck trips, composition and revenue; and (4) present the results with alerts to supervisors. Figure 4.1 shows this framework as implemented.

3. RESEARCH METHODOLOGY

3.1 Research Design

The study follows a quantitative, applied research design. A software artefact was designed and built, and its predictive component was evaluated empirically on held-out data. The evaluation answers RQ1 and RQ2 through error metrics computed per stream and per zone, and RQ3 by comparing errors with truck capacity. The work proceeded in five steps: data preparation, rainfall integration, model development, decision-support design and evaluation.

3.2 Data, Participants or Study Materials

The study uses operational records of the Kalutara Municipal Council. The dataset covers the period January 2023 to December 2025 and contains 628 zone-level weekly observations from four zones. No human participants were involved. The zones and their characteristics are given in Table 3.1.

Table 3.1: Collection zones in the study area

Zone	Area type	Population density	Dominant profile
Kalutara North	Town	Medium	Administrative offices, commercial markets, transport hub
Kalutara South	Beach	Low	Tourism resorts, coastal recreation, fishing harbors
Katukurunda 1	Muslim area	High	Dense residential clusters, community markets
Katukurunda 2	Town	Medium	Mixed residential and commercial corridor

Source: Kalutara Municipal Council records

Each record contains the year, week start and end dates, month, zone, area type, population density, week type (normal, holiday or festival), a special-event flag, 7-day cumulative rainfall in millimeters, the previous week's wet and dry tonnage, and two targets: wet waste tonnes and dry waste tonnes. The full schema is listed in Appendix A (Table A.1).

3.3 Experimental or Study Procedure

3.3.1 Automated rainfall retrieval

To avoid manual entry of weather data, the back-end resolves rainfall for a target week through a four-tier hierarchy. The first tier that returns a valid value is used.

1. Training dataset. If the week exists in the historical dataset, the recorded rainfall is used.
2. Open-Meteo forecast and past-days service. For weeks from about 92 days in the past to 16 days ahead, the 7-day rainfall is obtained from the Open-Meteo service [9] if all seven daily values are available.
3. Open-Meteo historical archive. For older dates, archived reanalysis data based on ERA5 [10] are retrieved.
4. Climatological average. For dates further than 16 days ahead, the monthly mean weekly rainfall derived from the training data is used (Appendix A, Table A.2).

3.3.2 Preprocessing and pipeline

Transformations are placed inside a scikit-learn pipeline [11] together with the estimator so that encoding is fitted only on training data and data leakage is avoided. The numerical features (month, rainfall, previous wet tonnage and previous dry tonnage) are passed through unchanged, since tree-based models do not need scaling. The categorical features (zone, area type, population density, week type and special event) are one-hot encoded, dropping the first category to avoid collinear indicators and ignoring categories not seen in training.

3.3.3 Model formulation

Wet waste is driven largely by food preparation, moisture and household behavior, whereas dry waste is driven by packaging, commercial supply chains and tourism. Two separate XGBoost regressors were therefore trained, one for each target. XGBoost minimizes the regularized objective in Equations 3.1 and 3.2 [7].

$$L(\theta) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (3.1)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_j w_j^2 \quad (3.2)$$

Here l is the loss between actual and predicted values, T is the number of leaves in a tree, w_j are the leaf weights, and γ and λ are regularization parameters. The configuration of the two models is given in Table 3.2. The wet-waste model uses shallower trees (depth 3) to limit fitting of noise, while the dry-waste model uses depth 5 so that it can represent interactions between area type, events and rainfall.

Table 3.2: Model configuration

Parameter	Wet waste model	Dry waste model
Algorithm	XGBRegressor	XGBRegressor
Learning rate (η)	0.05	0.05
Number of estimators	100	100
Maximum depth	3	5
Booster	gbtree	gbtree
Parallel jobs	-1 (all cores)	-1 (all cores)

3.3.4 Validation strategy

Random k-fold cross-validation mixes future and past observations and can overstate performance for time-ordered data [12]. A strict chronological split was therefore used. The first 80% of records (502 records, weeks from 9 January 2023 to 19 May 2025) formed the training set, and the final 20% (126 records, weeks from 19 May 2025 to 30 June 2026) formed the test set, which the models never saw during training.

3.4 Measures and Evaluation Metrics

Performance was measured with mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE) and the coefficient of determination (R^2) [13]. A municipal accuracy figure is reported as 100% minus MAPE, which is a convention used for readability and not a classification accuracy.

$$MAE = (1/n) \sum |y_i - \hat{y}_i| \quad (3.3)$$

$$RMSE = \sqrt{[(1/n) \sum (y_i - \hat{y}_i)^2]} \quad (3.4)$$

$$MAPE = (100\%/n) \sum |(y_i - \hat{y}_i) / y_i| \quad (3.5)$$

$$R^2 = 1 - \sum (y_i - \hat{y}_i)^2 / \sum (y_i - \bar{y})^2 \quad (3.6)$$

$$Predictive\ accuracy = 100\% - MAPE \quad (3.7)$$

MAE and RMSE are in tonnes and can be compared directly with truck capacity. RMSE penalizes large errors more than MAE. MAPE expresses error relative to the actual value, but it is sensitive to weeks with small tonnage.

3.5 Data Analysis

Metrics were computed on the 126 test records for each model, then separately for each zone, so that overall performance is not taken to represent every zone. The wet and dry results are compared to examine H1. Absolute errors are compared with the 5-tonne truck capacity to address RQ3. No baseline models were benchmarked in this study; this is recorded as a limitation in Section 7.5.

3.6 Ethics, Privacy, Security and Responsible AI

The waste records are aggregated tonnages per zone and week and contain no personal data about residents. The only personal data held by the application are user account credentials; passwords are stored as salted hashes and sessions use signed tokens with 24-hour expiry. The system is designed as decision support: forecasts are advisory and supervisors retain responsibility for dispatch decisions. Zone-level results are reported separately so that weaker performance in any zone is visible rather than hidden in an overall average.

3.7 Reproducibility and Research Artefacts

The research artefacts are the zonal dataset, the feature pipeline, the two trained models, the Flask back-end and the React front-end. Reproducibility is supported by the fixed chronological split, the listed hyper parameters (Table 3.2) and the software versions in Table 5.1.

4. PRODUCT AND SOLUTION DESIGN

4.1 Requirements

The requirements were derived from the problem statement and from the way KMC supervisors plan weekly collection. They are summarized in Table 4.1.

Table 4.1: Functional and non-functional requirements

ID	Requirement
FR1	Forecast weekly wet and dry waste tonnage for each zone.
FR2	Retrieve rainfall for the target week automatically.
FR3	Retrieve the previous week's tonnage per zone as a model input.
FR4	Calculate truck trips, trips per truck and fleet utilization.
FR5	Estimate material composition (organic, plastic, paper, metal).
FR6	Estimate weekly, monthly and annual recovery revenue in Sri Lankan rupees (LKR).
FR7	Run back testing on held-out records and display actual versus predicted values.
FR8	Generate rule-based operational recommendations and keep a history of predictions.
NFR1	Forecast errors should be small relative to the 5-tonne truck capacity.
NFR2	No manual weather entry; the system must still work when external services fail.
NFR3	Authenticated access, with passwords stored as salted hashes.
NFR4	A responsive web interface usable by non-technical supervisors.

4.2 Component Architecture within the Complete Solution

The component is a three-tier web application. Data enter from the municipal dataset and the Open-Meteo service, pass through the feature pipeline and the two XGBoost models, and then feed four decision engines (fleet logistics, composition, revenue and recommendations). Results are returned to a React dashboard and stored in MongoDB as an audit trail. Figure 4.1 shows the architecture.

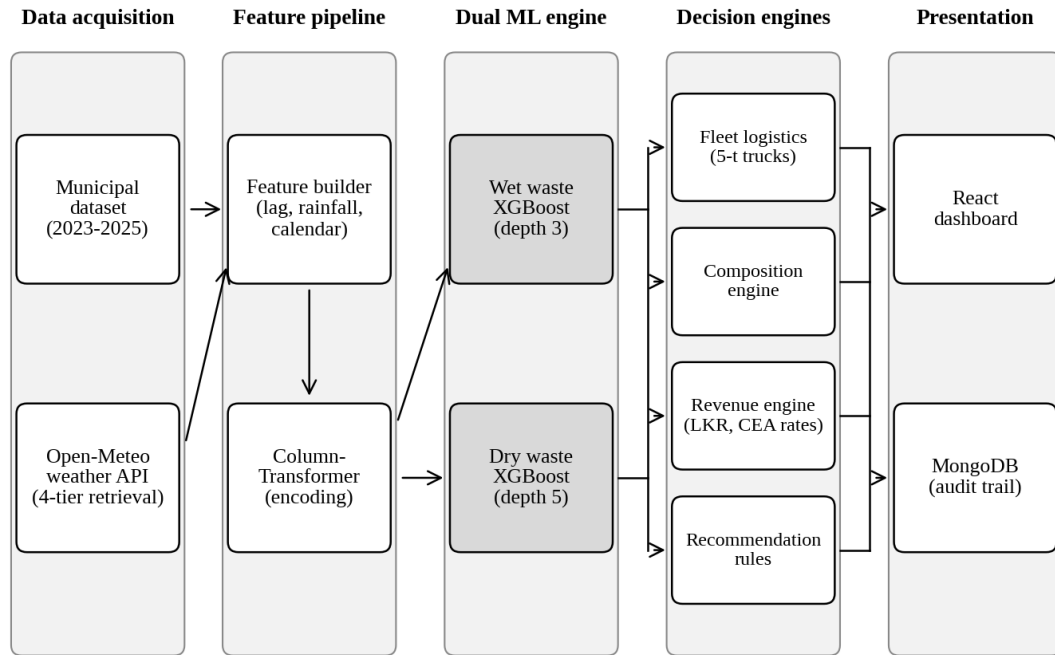


Figure 4.1: Architecture of the waste prediction and resource monetization component

The component exposes its predicted tonnage per zone through REST endpoints. This is the interface through which the smart-bin components (Components 1 and 2) and the route optimizer (Component 4) can use the forecasts in future integration, as discussed in Section 9.4.

4.3 Design Alternatives and Decisions

Key design decisions and the alternatives considered are listed in Table 4.2. The choices of algorithm were made on the basis of the literature reviewed in Chapter 2; competing algorithms were not benchmarked experimentally in this study.

Table 4.2: Main design decisions

Decision	Alternative	Rationale
XGBoost regression	ARIMA, neural network, random forest	Non-linear, regularized, suited to small tabular data (Section 2.2)
Separate wet and dry models	One multi-output model	Different drivers and noise levels; allows a different tree depth for each
Pass-through numeric features	Scaling	Tree splits are scale-invariant; keeps units interpretable

Decision	Alternative	Rationale
One-hot encoding with unseen-category handling	Label encoding	Avoids false ordering; tolerates new categories at inference
Chronological 80/20 split	Random k-fold	Prevents look-ahead leakage in time-ordered data
Four-tier rainfall retrieval	Manual entry	Removes manual input and degrades gracefully when services fail
Rule-based recommendations	Learned recommender	Transparent and easy for officers to audit

4.4 Detailed Design

4.4.1 Fleet requirement planning

For each zone z , the weekly load is the sum of the predicted wet and dry tonnage, $W_z = \hat{y}_{wet,z} + \hat{y}_{dry,z}$. With truck capacity $C = 5.0$ tonnes, the trips required for a zone and for the whole council are given by Equations 4.1 and 4.2. For a fleet of F vehicles ($F = 3$), the trips per truck and the overall utilisation follow from Equations 4.3 and 4.4.

$$L_z = \lceil W_z / C \rceil \quad (4.1)$$

$$L_{total} = \lceil \sum W_z / C \rceil \quad (4.2)$$

$$T_{truck} = \lceil L_{total} / F \rceil \quad (4.3)$$

$$U = \sum W_z / (L_{total} \times C) \times 100\% \quad (4.4)$$

4.4.2 Waste composition estimation

The material composition of a zone is estimated by applying percentage shares that depend on its area type (Table 4.3). These shares are benchmark values for coastal Sri Lankan municipal waste. The council-wide tonnage of material m is calculated with Equation 4.5, where $\kappa(z,m)$ is the share of material m in zone z .

$$M_m = \sum_z (W_z \times \kappa(z,m) / 100) \quad (4.5)$$

Table 4.3: Assumed waste composition by area type (%)

Area type	Organic	Plastic	Paper	Metal
Town	55	20	15	10
Beach	40	30	18	12

Area type	Organic	Plastic	Paper	Metal
Muslim area	62	18	12	8

4.4.3 Circular economy revenue model

Predicted wet and dry tonnage is allocated to recovery routes using the shares, yields and benchmark prices in Table 4.4. The prices are based on CEA and Mihisaru national standards and on rates quoted by registered private recyclers. Weekly revenue is the sum over routes of tonnage multiplied by rate (Equation 4.6) and is scaled to monthly and annual values with Equations 4.7 and 4.8.

Table 4.4: Resource recovery routes and benchmark rates

Stream	Route	Share of stream	Yield	Benchmark rate
Wet	Aerobic compost (Mihisaru standard)	70%	35%	15 LKR/kg of cured compost
Wet	Anaerobic digestion (energy offset)	20%	–	12,000 LKR/tonne
Wet	Animal feed / soil conditioner	10%	–	10,000 LKR/tonne
Dry	Plastics (PET, HDPE, PP)	30%	85%	75 LKR/kg
Dry	Cardboard and kraft paper	35%	80%	32 LKR/kg
Dry	Glass cullet	15%	95%	16 LKR/kg
Dry	Scrap metals and aluminium cans	12%	90%	90 LKR/kg
Dry	Non-recyclable residue	8%	–	0 (landfill / refuse-derived fuel, RDF)

Source: CEA and Mihisaru benchmarks as configured in the system

$$Revenue_week = \sum_k (Tonnage_k \times Rate_k) \quad (4.6)$$

$$Revenue_month = Revenue_week \times 4.33 \quad (4.7)$$

$$Revenue_year = Revenue_week \times 52 \quad (4.8)$$

4.4.4 Recommendation rules

A rule-based module compares predicted values with thresholds and raises alerts, for example for unusually high weekly generation, for compost-facility capacity planning and for prioritizing beach clean-ups. Rules were chosen over a learned model so that supervisors can see exactly why an alert was raised.

4.5 Standards and Professional Practice

The design follows the CEA and Mihisaru benchmarks for recovery values, uses Representational State Transfer (REST) conventions for the interface and token-based authentication for access control, and keeps secrets in environment variables rather than in source code. Chronological validation and per-zone reporting follow accepted practice for time-ordered data and responsible reporting of model performance.

5. IMPLEMENTATION, INTEGRATION AND TESTING

5.1 Development Environment and Technology Stack

The back-end is a Python service and the front-end is a single-page web application. Table 5.1 lists the main technologies.

Table 5.1: Technology stack

Layer	Technologies
Back-end	Python 3.13, Flask 3.1.1, flask-cors, python-dotenv
Machine learning	XGBoost, scikit-learn 1.6.1 (Pipeline, ColumnTransformer)
Database	MongoDB Atlas with PyMongo 4.17.0
Authentication	PyJWT, werkzeug.security (salted password hashing)
Front-end	React 18, Vite, Tailwind CSS, Framer Motion, Lucide Icons, Recharts
External data	Open-Meteo forecast and historical archive APIs

5.2 Component Implementation

5.2.1 Back-end services

The Flask application loads the two trained pipelines at start-up and exposes the REST endpoints in Table 5.2. A prediction request supplies the week type and, optionally, the week start date. The service then looks up rainfall, retrieves the previous week's tonnage for each zone, runs both pipelines, applies the fleet, composition and revenue calculations, and stores the result in the predictions collection of MongoDB. Access is controlled by bearer tokens signed with PyJWT and valid for 24 hours.

Table 5.2: Main REST endpoints

Endpoint	Purpose
POST /api/predict	Run forecasts, truck planning, composition and revenue for a target week
GET /api/rainfall-forecast	Return 7-day cumulative rainfall for a week using the four-tier hierarchy
GET /api/previous-week	Return lag-1 wet and dry tonnage per zone
GET /api/backtest	Return actual versus predicted values and metrics for held-out records
GET /api/predictions	Return paginated prediction history
POST /api/auth/login, /api/auth/register	Authenticate and register users

5.2.2 Front-end dashboard

The dashboard communicates with the back-end through a central API module and presents results with Recharts visualizations. Its main components are:

- PredictionInputPanel – week selection and automatic rainfall feedback;
- PredictionResults – wet, dry and total tonnage cards and daily breakdowns;
- RevenueEstimation – weekly, monthly and annual revenue views and material breakdown;
- ModelValidation – actual versus predicted overlays and scatter plots with zone filters;
- Recommendations – rule-based alerts;
- PredictionHistory – audit table of earlier runs.

5.3 Integration with the Other Components

Internally, the front-end, back-end, models, database and weather service are integrated through the endpoints above. The rainfall service is the only external dependency during prediction, and the four-tier design ensures that a value is always returned, even when the live service is unavailable. Integration with the smart-bin components and the route optimizer is not yet implemented; the zone-level forecasts are available through the API for that purpose.

5.4 Verification and Software Testing

Table 5.3 lists the verification measures applied to the main risks. The model-related controls are evidenced by the results in Chapter 6.

Table 5.3: Verification measures

Risk	Control implemented	Evidence
Data leakage	Encoding inside the pipeline; chronological split	Held-out results in Table 6.1
Unseen categories at inference	One-hot encoder set to ignore unknown values	Pipeline configuration (Section 3.3.2)
Weather service failure	Four-tier fallback including climatological means	Endpoint behavior (Section 5.3)
Unauthorized access	Signed tokens; hashed passwords	Authentication endpoints (Table 5.2)

Risk	Control implemented	Evidence
Misleading performance claims	Backtesting endpoint reproducing test metrics	Dashboard validation view

5.5 Deployment and Operation

In operation, a supervisor logs in, selects the coming week and receives forecasts without entering weather data. Configuration secrets are injected through environment variables, the database is hosted on MongoDB Atlas, and every prediction is logged for audit. Packaging the inference pipeline for deployment on a council's own server is proposed as future work.

6. RESULTS AND VALIDATION

6.1 Research Results

Table 6.1 shows the performance of the two models on the 126 held-out records. The dry-waste model explained 94.92% of the variance ($R^2 = 0.9492$) with an MAE of 1.521 tonnes and a predictive accuracy of 87.67%. The wet-waste model achieved $R^2 = 0.8035$, an MAE of 3.275 tonnes and a predictive accuracy of 80.85%.

Table 6.1: Overall model performance on the test set (n = 126)

Model	MAE (t)	RMSE (t)	MAPE (%)	Accuracy (%)	R^2
Wet waste	3.275	4.826	19.15	80.85	0.8035
Dry waste	1.521	2.273	12.33	87.67	0.9492

Table 6.2 breaks the results down by zone. Dry-waste accuracy ranged from 85.35% (Kalutara South) to 89.00% (Katukurunda 1), and dry-waste R^2 was above 0.90 in every zone. Wet-waste accuracy ranged from 79.43% to 82.83%, but zone-level wet-waste R^2 was much lower, between 0.2994 and 0.4485. Figure 6.1 compares accuracy by zone and stream.

Table 6.2: Zone-level performance on the test set

Zone	Stream	Mean actual (t)	MAE (t)	RMSE (t)	MAPE (%)	Acc. (%)	R^2
Kalutara North	Wet	14.28	2.721	3.779	20.34	79.66	0.3631
	Dry	10.42	1.197	1.590	11.86	88.14	0.9392
Kalutara South	Wet	10.63	2.072	2.915	20.57	79.43	0.2994
	Dry	7.64	1.091	1.533	14.65	85.35	0.9044
Katukurunda 1	Wet	28.52	5.221	7.148	18.61	81.39	0.3918
	Dry	20.47	2.193	3.239	11.00	89.00	0.9383
Katukurunda 2	Wet	18.15	3.033	4.306	17.17	82.83	0.4485
	Dry	13.56	1.580	2.263	11.89	88.11	0.9227

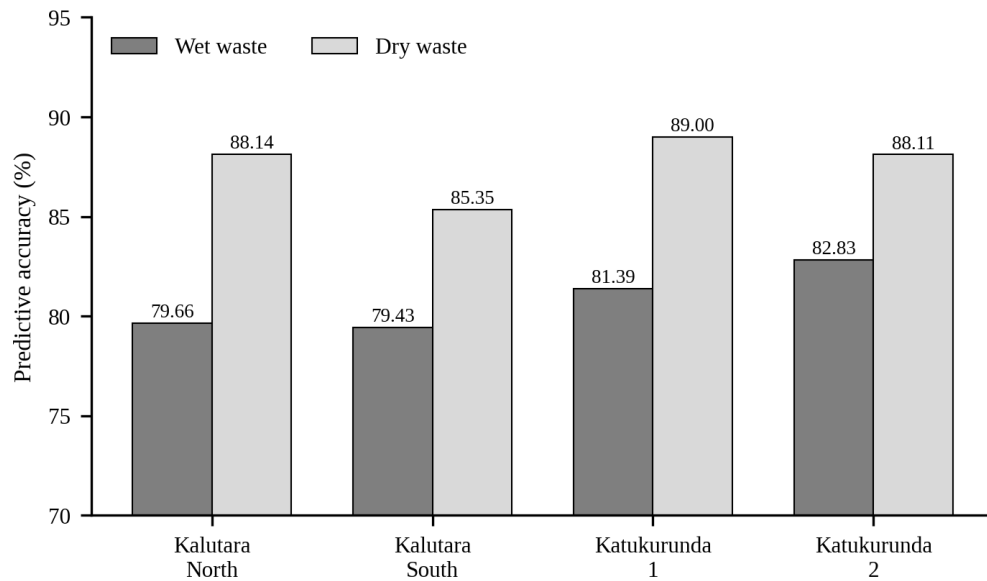


Figure 6.1: Predictive accuracy by zone and waste stream on the test set
Source: values from Table 6.2

6.2 Product and Technical Results

The completed system produces, for any selected week, zone-level and council-wide wet, dry and total tonnage; the number of 5-tonne truck trips, trips per truck and utilization; the material composition; revenue estimates for weekly, monthly and annual horizons; and a list of recommendations. The back testing view allows supervisors to compare the forecasts with actual records for any zone and thereby judge the reliability of the model for themselves.

6.3 User or Stakeholder Validation

A structured evaluation with KMC staff is not reported in this thesis; validation here is technical and based on held-out data. Stakeholder evaluation is included in the future work in Section 9.4.

6.4 Integration Results

The back-end, models, database, weather service and dashboard work together end to end: a single request to the prediction endpoint returns forecasts, truck planning, composition and revenue and logs the run. The four-tier rainfall hierarchy supplied a rainfall value for any requested week, using recorded data, live or archived services, or the monthly climatological averages as appropriate.

6.5 Summary of Findings

- Dry waste was forecast more accurately than wet waste (accuracy 87.67% against 80.85%; R^2 0.9492 against 0.8035), which is consistent with H1.
- Dry-waste R^2 exceeded 0.90 in all four zones, whereas wet-waste R^2 within individual zones was between 0.30 and 0.45.
- Average absolute errors per zone-week (1.521 t for dry and 3.275 t for wet waste) were below the capacity of one 5-tonne truck.
- The forecasts were successfully converted into truck trips, composition and revenue estimates in a working dashboard.

7. DISCUSSION

7.1 Interpretation of Findings

The results support the working hypothesis that dry waste is more predictable than wet waste. Dry-waste generation is closely tied to packaging, commercial activity and tourism, which are relatively stable from week to week, and the previous week's dry tonnage is therefore a strong predictor. Wet waste is also influenced by moisture, household composting and food preparation, which vary more and are only partly captured by rainfall and calendar features.

The zone-level results qualify the headline figure for wet waste. The pooled wet-waste R^2 of 0.8035 is much higher than the zone-level values of 0.30 to 0.45. A likely reason is that the pooled figure includes the large differences between zones: mean weekly wet waste ranges from 10.63 tonnes in Kalutara South to 28.52 tonnes in Katukurunda 1, so separating the zones already explains a large part of the total variance. Within a zone, the model explains much less of the week-to-week variation. Reporting both pooled and per-zone results is therefore important for an honest interpretation.

In operational terms, the average absolute errors of 1.521 tonnes (dry) and 3.275 tonnes (wet) per zone-week are below the capacity of a single 5-tonne truck. This suggests that forecasts can inform decisions such as how many trips to schedule, although errors are larger in the highest-volume zone (Katukurunda 1, wet MAE 5.221 tonnes), where a dispatch decision could be off by about one truck.

7.2 Comparison with Literature Baselines and Existing Solutions

This study did not implement ARIMA, random forest or neural network baselines on the KMC data, so it cannot claim that XGBoost outperforms them here. The choice of XGBoost rests on the literature [7], [8]. Comparison of the absolute accuracy figures with published studies is also not meaningful, because datasets, forecast horizons and spatial scales differ. Relative to current KMC practice, which has no quantitative forecast, the system adds a measurable and testable forecast. Relative to sensor-only smart-bin systems, it adds a planning horizon and links tonnage to fleet and revenue decisions.

7.3 Research Contribution

- A validated weekly, zone-level forecasting model for wet and dry waste using real municipal records from a coastal Sri Lankan council.
- A rainfall retrieval design that makes the model usable without manual weather entry and that degrades gracefully to climatological values.
- Evidence that wet and dry streams differ in predictability and that zone-level reporting is necessary to interpret pooled results.

7.4 Product Contribution

The product contribution is a working decision-support dashboard that turns forecasts into truck trips, composition estimates, revenue estimates and alerts, with backtesting and an audit trail. It gives KMC a tool that can be used in weekly planning and can be extended with the other components of the group project.

7.5 Limitations and Threats to Validity

1. Data scope. The models are trained on about three years of records from one council and four zones. Seasonal patterns are represented by only a few cycles, and results may not transfer to other councils without retraining.
2. Dependence on lagged input. The forecast needs the previous week's tonnage. The relative contribution of the lag, rainfall and calendar features was not isolated; an ablation or feature-importance analysis would show it.
3. No baselines or hyperparameter search. Alternative algorithms were not benchmarked and the configuration in Table 3.2 was not tuned through a reported search.
4. Weather uncertainty. Forecast rainfall becomes less reliable beyond 7 to 10 days, and the climatological fallback ignores the actual conditions of a given week.
5. Informal sector. Informal waste pickers remove some high-value recyclables before collection, which adds noise to the dry-waste records and means that revenue estimates may be optimistic.

6. Fixed assumptions. Composition shares, recovery yields and prices are benchmark values and not measured or live. Metrics such as MAPE are also sensitive to weeks with low tonnage.
7. Limited stakeholder validation. No structured user evaluation has yet been carried out.

7.6 Implications

For KMC, the results indicate that forecast-based planning of crews and vehicles is feasible with data that the council already records. For other local authorities, the approach could be replicated by recording weekly tonnage by zone and retraining the models. For research, the findings show the value of combining pooled and per-zone evaluation when reporting waste forecasts.

8. COMMERCIALISATION AND SUSTAINABILITY

8.1 Value Proposition and Target Users

The primary users are municipal health inspectors and logistics supervisors, who gain advance knowledge of weekly tonnage and the number of truck trips needed. Secondary users are compost plant and material recovery facility managers, who can plan labor and storage, and recyclers, who can plan purchases from predictable inbound volumes. The main value to a council is lower fuel and operator cost from better-loaded trips, fewer overflows, and a basis for earning revenue from segregated materials.

8.2 Market and Competitive Position

The potential market consists of local authorities in Sri Lanka and similar developing-country councils that collect waste by fixed schedules and keep weekly tonnage records. Existing smart-bin products concentrate on sensing the fill level of containers. The distinguishing feature of this component is a localized forecast, with automatic weather input, linked to fleet and revenue decisions.

8.3 Business or Sustainability Model

The system can be provided to councils as a subscription service or as a publicly funded tool maintained by a technology partner. Revenue from recovered materials accrues to the council and helps to fund the service. Table 8.1 illustrates the value of the dry-waste recovery routes, calculated from the shares, yields and rates in Table 4.4.

Table 8.1: Illustrative revenue per tonne of dry waste

Recovery route	Share $\times$ yield	Rate (LKR/t)	Revenue per tonne of dry waste (LKR)
Plastics	0.30×0.85	75,000	19,125
Cardboard and paper	0.35×0.80	32,000	8,960
Glass cullet	0.15×0.95	16,000	2,280
Scrap metals	0.12×0.90	90,000	9,720
Residue	0.08	0	0
Total			40,085

Source: calculation from Table 4.4

Applying this value to the mean combined weekly dry-waste tonnage of the four zones in the test period (52.09 tonnes) gives a gross figure of about LKR 2.09 million per week, or about LKR 108.6 million per year using Equation 4.8. This is an upper-bound illustration. It assumes that all dry waste reaches the stated routes at benchmark yields and prices, and it excludes sorting, processing, transport and labor costs and the effect of informal collection. A cost model and a sensitivity analysis are needed before the figure is used for investment decisions.

8.4 Scalability, Intellectual Property and Compliance

The architecture is scalable because inference is lightweight and the models can be retrained for a new council from its own weekly records. The solution is built on open-source libraries, and ownership of the dataset rests with KMC. User accounts hold personal data, so the system should comply with the Personal Data Protection Act, No. 9 of 2022 of Sri Lanka [14].

8.5 Contribution to the Sustainable Development Goals

The component contributes to the United Nations Sustainable Development Goals (SDGs) [15]. It supports SDG 11 (sustainable cities and communities) through better municipal waste collection, SDG 12 (responsible consumption and production) by encouraging recovery of compost and recyclables, and SDG 13 (climate action) through avoided fuel use and the diversion of organic waste from open disposal.

9. CONCLUSION AND FUTURE DIRECTIONS

9.1 Conclusion

This research developed and evaluated a machine learning platform that forecasts weekly wet and dry municipal waste for the zones of the Kalutara Municipal Council and converts the forecasts into fleet and revenue information. On a chronological hold-out test, the dry-waste model reached $R^2 = 0.9492$ and 87.67% predictive accuracy, and the wet-waste model reached $R^2 = 0.8035$ and 80.85% accuracy. Per-zone results showed that wet-waste forecasting within a zone is considerably less accurate than the pooled figure suggests. The automated rainfall retrieval and the decision engines allow the forecasts to be used in planning without manual data entry.

9.2 Achievement of Objectives

Table 9.1: Achievement of objectives

Objective	Evidence	Status
1. Separate wet and dry models	Two XGBoost pipelines; Tables 3.2, 6.1	Achieved
2. Automated rainfall retrieval	Four-tier hierarchy; Section 3.3.1	Achieved
3. Composition estimation	Table 4.3; Equation 4.5	Achieved (benchmark shares)
4. Truck logistics	Equations 4.1–4.4; dashboard	Achieved
5. Revenue estimation	Table 4.4; Table 8.1	Achieved (benchmark prices)
6. Interactive dashboard	Section 5.2.2	Achieved; user evaluation pending

9.3 Individual Research and Product Contributions

The research contribution is the validated forecasting approach and the finding that wet and dry streams differ in predictability. The product contribution is the integrated back-end and dashboard that applies the forecasts to dispatch, composition and revenue decisions for KMC.

9.4 Future Directions

- Benchmark ARIMA, random forest and neural network baselines, and analyze feature importance and hyperparameter setting.
- Extend the dataset to more years and more councils and add measured composition and sorting data.
- Integrate with the smart-bin sensors (Components 1 and 2) and supply forecasts to the route optimizer (Component 4).
- Develop a cost and sensitivity model for the revenue estimates and include informal-sector effects.
- Carry out a structured evaluation with KMC staff and package the inference service for deployment on council servers.

REFERENCES

- [1] S. Kaza, L. Yao, P. Bhada-Tata, and F. Van Woerden, *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*. Washington, DC, USA: World Bank, 2018.
- [2] Central Environmental Authority, Sri Lanka. “Central Environmental Authority.” [Online]. Available: <https://www.cea.lk> [Accessed: 1 Oct. 2026].
- [3] A. Abbasi and A. El Hanandeh, “Forecasting municipal solid waste generation using artificial intelligence modelling approaches,” *Waste Manag.*, vol. 56, pp. 13–22, 2016.
- [4] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: Wiley, 2015.
- [5] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [6] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [7] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794.
- [8] L. Grinsztajn, E. Oyallon, and G. Varoquaux, “Why do tree-based models still outperform deep learning on typical tabular data?” in *Adv. Neural Inf. Process. Syst. 35 (NeurIPS 2022), Datasets and Benchmarks Track*, 2022.
- [9] Open-Meteo, “Open-Meteo weather forecast and historical weather APIs.” [Online]. Available: <https://open-meteo.com> [Accessed: 1 Oct. 2026].
- [10] H. Hersbach et al., “The ERA5 global reanalysis,” *Q. J. R. Meteorol. Soc.*, vol. 146, no. 730, pp. 1999–2049, 2020.
- [11] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [12] C. Bergmeir and J. M. Benítez, “On the use of cross-validation for time series predictor evaluation,” *Inf. Sci.*, vol. 191, pp. 192–213, 2012.
- [13] R. J. Hyndman and A. B. Koehler, “Another look at measures of forecast accuracy,” *Int. J. Forecast.*, vol. 22, no. 4, pp. 679–688, 2006.
- [14] Parliament of the Democratic Socialist Republic of Sri Lanka, *Personal Data Protection Act, No. 9 of 2022*. Colombo, Sri Lanka, 2022.

- [15] United Nations General Assembly, “Transforming our world: The 2030 Agenda for Sustainable Development,” Resolution A/RES/70/1, 2015.

GLOSSARY

Backtesting – Evaluating a model on past records that it did not see during training.

Circular economy – An approach in which waste is treated as a resource and recovered for reuse.

Compactor truck – A refuse vehicle that compresses waste; the standard capacity used here is 5 tonnes.

Gradient boosting – An ensemble method that adds decision trees one at a time to reduce the remaining error.

Lag feature – The value of a variable from an earlier period, here the previous week's waste tonnage.

Mihisaru – A Sri Lankan national programme and standard associated with composting and waste recovery.

Poya day – A monthly full-moon public holiday in Sri Lanka.

Wet and dry waste – Biodegradable (kitchen and garden) waste, and recyclable or residual non-biodegradable waste.

APPENDICES

Appendix A: Extended Results and Supplementary Specifications

Table A.1 lists the dataset schema referred to in Section 3.2, and Table A.2 gives the monthly climatological rainfall used as the fourth-tier fallback in Section 3.3.1.

Table A.1: Dataset schema

Field	Type	Description
year	Integer	Calendar year (2023–2025)
week_start_date, week_end_date	Date	Boundaries of the calendar week
month	Integer	Month index (1–12)
zone	Nominal	Collection zone identifier
area_type	Nominal	Town, beach or Muslim area
population_density	Nominal	Low, medium or high
week_type	Nominal	Normal, holiday or festival
special_event	Binary	Yes or no flag for cultural or religious festivals
rainfall_mm	Numeric	7-day cumulative rainfall (mm)
previous_wet_tons	Numeric	Wet waste in the previous week (t)
previous_dry_tons	Numeric	Dry waste in the previous week (t)
wet_waste_tons	Numeric (target)	Wet waste collected (t)
dry_waste_tons	Numeric (target)	Dry waste collected (t)

Table A.2: Climatological mean weekly rainfall used as fallback (mm)

Month	mm	Month	mm	Month	mm
January	11.5	May	77.7	September	62.0
February	10.5	June	63.8	October	68.0
March	25.0	July	47.8	November	53.9
April	50.3	August	38.7	December	46.7

Appendix B: Academic Integrity and AI Use Disclosure

This thesis is based on my own research, system development, implementation, testing, analysis, and results.

ChatGPT and Claude were used only as supporting tools to improve language clarity, sentence structure, and the organization of selected thesis sections. They were not used

to create research data, alter experimental results, fabricate references, or make research decisions.

All AI-assisted content was reviewed and verified against my original research work, implementation, results, and sources. I take full responsibility for the accuracy, originality, and conclusions presented in this thesis.

Appendix C: System Implementation Evidence

C.1 Dataset Evidence

The following screenshots provide samples of the dataset used for model development and evaluation.

Training Data

2025 වර්ෂයේ එක්සත් කරන ලද කළමනාකරණ දත්ත

මාසය	දිනක කසල (ප්‍රමාණය) - kg	තොරතුරු කසල (ප්‍රමාණය) - kg
ජනවාරි	376661	185560
පෙබරවාරි	327742	132173
මාර්තු	445019	122634
අප්‍රේල්	425674	131790
මැයි	461161	135979
ජූනි	487436	118900
ජූලි	429339	155540
අගෝස්තු	363719	185270
සැප්තැම්බර්	440537	143418
ඔක්තෝබර්	543732	543732
නොවැම්බර්	182185	403750
දෙසැම්බර්	421209	145265
මුළු එකතුව		

W. M. NISSANIGA VILAKKON
Administrative Officer
Postal Add No. 20-1100 &
Public Health Inspector
KALUTARA - 04

Training Data

2023 වර්ෂයේ එක්සත් කරන ලද කළමනාකරණ දත්ත

මාසය	දිනක කසල (ප්‍රමාණය) - kg	තොරතුරු කසල (ප්‍රමාණය) - kg
ජනවාරි	345286	176971
පෙබරවාරි	354829	140260
මාර්තු	432232	136900
අප්‍රේල්	445971	173754
මැයි	432497	142678
ජූනි	454740	130502
ජූලි	442000	148450
අගෝස්තු	298651	187885
සැප්තැම්බර්	474837	167826
ඔක්තෝබර්	501304	457620
නොවැම්බර්	370284	492531
දෙසැම්බර්	442796	164733
මුළු එකතුව		

Training Data

2024 වර්ෂයේ එක්සත් කරන ලද කළමනාකරණ දත්ත

මාසය	දිනක කසල (ප්‍රමාණය) - kg	තොරතුරු කසල (ප්‍රමාණය) - kg
ජනවාරි	359285	183940
පෙබරවාරි	332628	141284
මාර්තු	447637	136543
අප්‍රේල්	435589	151222
මැයි	452598	141325
ජූනි	468643	121563
ජූලි	445432	156427
අගෝස්තු	349783	173456
සැප්තැම්බර්	450864	153775
ඔක්තෝබර්	498891	489534
නොවැම්බර්	378495	507523
දෙසැම්බර්	465123	153510
මුළු එකතුව		

4 zones Kalutara

- > Kalutara North - 20% of waste → Medium population (town area)
- > Kalutara South - 15% of waste → low population (beach area)
- > Kankununda 1 - 40% of waste → high population (muslim area)
- > Kankununda 2 - 25% of waste → medium population (town area)

4 Supervisors

Test Data (Unseen Data)

2026 වර්ෂයේ එක්සත් කරන ලද කළමනාකරණ දත්ත

මාසය	දිනක කසල (ප්‍රමාණය) - kg	තොරතුරු කසල (ප්‍රමාණය) - kg
ජනවාරි	470795	188766
පෙබරවාරි	425744	147214
මාර්තු	396943	139143
අප්‍රේල්	521329	179508
මැයි	436269	143531
ජූනි	430621	144043
මුළු එකතුව		

Figure C.1: Sample records from the municipal waste dataset used for model development

C.2 Machine Learning Model Evidence

The following screenshots provide implementation evidence of the machine learning model development, execution, and evaluation process.

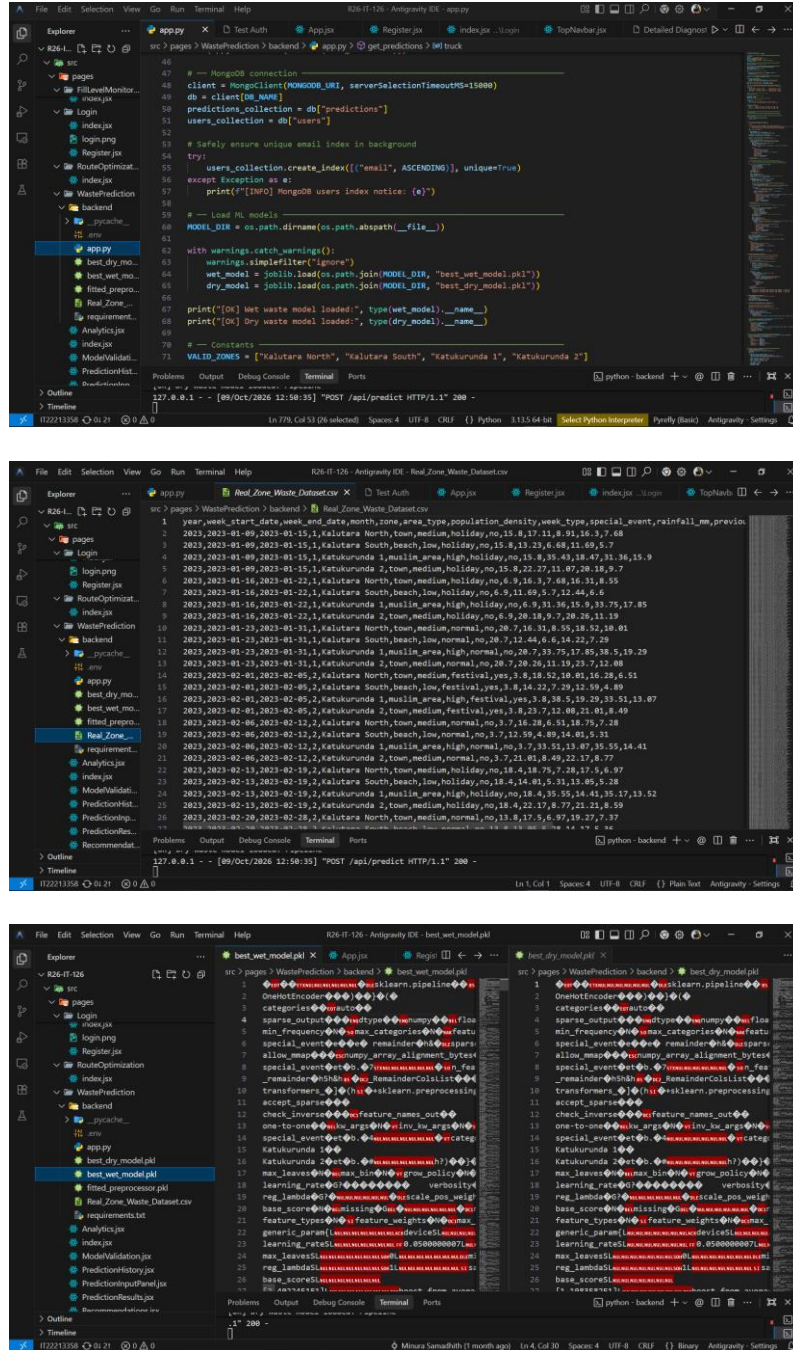


Figure C.2: Implementation of the machine learning prediction pipeline

C.3 Dashboard Evidence

The following screenshots demonstrate the implemented waste prediction and decision-support dashboard.

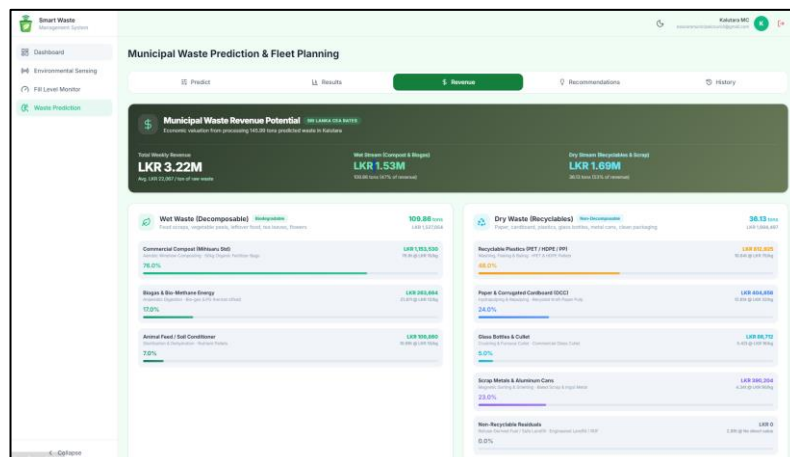
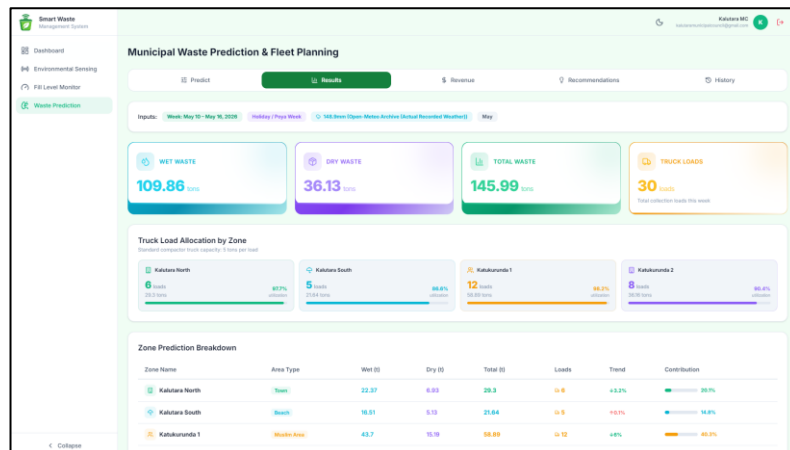
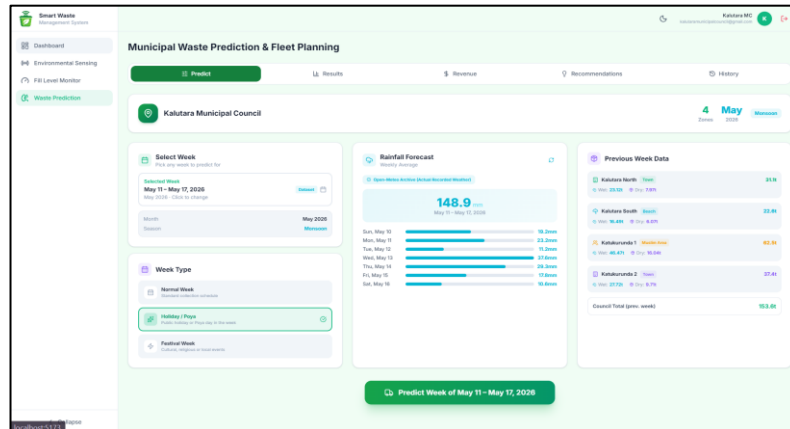


Figure C.3: Main waste prediction dashboard