

Smart Waste Management System

Project ID: R26-IT-126

Project Proposal Report

B.Sc. (Hons) Degree in Information Technology Specialized in
Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology Sri Lanka

March 2026

Waste Composition Prediction (ML-Based)

Project ID: R26-IT-126

Project Proposal Report

Jayawardhane K.G.M.S. – IT22213358

B.Sc. (Hons) Degree in Information Technology Specialized in
Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology Sri Lanka

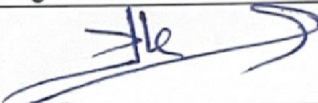
March 2026

DECLARATION

I declare that this is my own work, and this proposal does not knowingly incorporate any material previously submitted for a degree or diploma at any other university or higher education institution, nor does it contain any material that has been previously published or written by another person apart from where proper acknowledgment is given in the text.

Name	Student ID	Signature
Jayawardhane K.G.M.S.	IT22213358	

The above candidate is conducting research for their undergraduate Dissertation under my supervision.

Name	Date	Signature
Udisha D	20/03/26	

ABSTRACT

Rapid urbanization and increasing consumption have significantly intensified waste management challenges, particularly in developing countries such as Sri Lanka. Conventional waste management systems primarily rely on fixed collection schedules and reactive operational strategies where waste is collected only after accumulation occurs. Although modern smart waste management approaches have introduced technologies such as Internet of Things (IoT) sensors and real-time bin monitoring, most of these systems focus on detecting waste levels after disposal. As a result, waste management authorities and recycling facilities often lack prior knowledge of the types and proportions of waste expected during upcoming collection cycles. This uncertainty may lead to inefficient recycling preparation, poor resource allocation, and increased operational costs.

A significant research gap exists in predicting waste composition before the collection process begins. Most existing studies emphasize real-time monitoring of bin conditions or optimization of collection routes, while limited research has explored forecasting waste composition using predictive analytics. Addressing this limitation can support proactive planning and improve recycling efficiency.

This research proposes a Machine Learning–based Waste Composition Forecasting System that predicts the expected distribution of waste categories for the next collection cycle. The model analyzes historical waste collection data together with contextual factors such as weather conditions, calendar information, and location characteristics to identify patterns influencing waste generation. Machine learning algorithms including Random Forest and XGBoost will be used to estimate the percentage composition of waste types such as organic waste, plastic, paper, and metal.

The expected outcome is a predictive decision-support tool that assists waste management organizations in planning recycling operations in advance. The proposed system can contribute to improve recycling efficiency, optimized resource utilization, and more sustainable waste management practices.

Keywords: Waste Composition Prediction, Smart Waste Management, Machine Learning, Waste Forecasting, Predictive Analytics.

ACKNOWLEDGEMENT

I would like to begin by expressing my sincere appreciation to my family, especially my parents, for their endless support, patience, and encouragement throughout my academic journey. Their constant motivation and belief in my abilities have been a strong source of inspiration that helped me overcome many challenges during my studies.

My heartfelt gratitude goes to my supervisors and lecturers for their valuable guidance, professional advice, and continuous encouragement during the preparation of this research proposal. Their expertise and thoughtful feedback greatly contributed to improving the quality and direction of this research.

I am also thankful to my fellow colleagues and friends who provided support, shared ideas, and offered constructive discussions during the course of this study. Their cooperation and positive contributions made this research experience both productive and rewarding.

Lastly, I would like to acknowledge all individuals and organizations who willingly shared their knowledge and insights, which helped in understanding the practical aspects related to this research area. Their support and willingness to assist were greatly appreciated.

TABLE OF CONTENTS

DECLARATION	3
ABSTRACT	4
ACKNOWLEDGEMENT	5
LIST OF TABLES	8
LIST OF FIGURES	9
LIST OF ABBREVIATIONS	10
LIST OF APPENDICES	11
1.INTRODUCTION	12
1.1 Background and Context	12
1.2 Literature Review	13
1.3 Research Gap	15
2.OBJECTIVE.....	17
2.1 Main Objective	17
2.2 Specific Objectives	17
3.METHODOLOGY	18
3.1 Research Approach.....	18
3.2 Technical Methods and Algorithms	18
3.3 Tools and Platforms	19
3.4 Data Collection and Ethical Considerations	19
3.5 Validation Strategy and Performance Metrics	20
4.HIGH-LEVEL SYSTEM ARCHITECTURE	20
4.1 Overview of the Proposed System	20
4.2 System Components	21
4.3 Data Flow Explanation.....	22
4.5 Scalability and Security Considerations.....	23
5.USER REQUIREMENTS	23
5.1 Stakeholder Identification.....	23
5.2 Functional Requirements	24
5.3 Non-Functional Requirements.....	24
6.COMMERCIALIZATION PLAN	25
6.1 Target Market and Customer Persona.....	25
6.2 Value Proposition	26
6.3 Revenue Model	26
6.4 Cost Estimation.....	27

6.5 Pricing Strategy.....	27
6.6 Competitive Advantage	27
6.7 Intellectual Property Considerations.....	28
7.BUDGET AND JUSTIFICATION	28
8.WORK BREAKDOWN STRUCTURE (WBS)	31
9.GANTT CHART	33
10.REFERENCES LIST	34
11.APPENDICES	36

LIST OF TABLES

Table 1: Research Gap Summary table	16
Table 2: Technologies and Tools that will be used.....	19
Table 3: Target Market Segments	25
Table 4: Customer Persona	26
Table 5: Potential Revenue Streams	27
Table 6: Comparison with Existing Solutions	28
Table 7: Estimated Hardware Costs	29
Table 8: Software Licenses Costs	29
Table 9: Cloud Services Costs	30
Table 10: Estimated Human Effort hours	30
Table 11: Estimated Human Effort hours.....	32
Table 12: Workload Distribution.....	32

LIST OF FIGURES

Figure 1: High-Level Architecture Diagram of the System.....	21
Figure 2: Component Architecture Diagram.....	21
Figure 3: Work Breakdown Structure of the overall project.....	31
Figure 4: Gantt chart	33

LIST OF ABBREVIATIONS

Abbreviation	Meaning
AI	Artificial Intelligence
API	Application Programming Interface
AWS	Amazon Web Services
DSR	Design Science Research
IoT	Internet of Things
MAE	Mean Absolute Error
ML	Machine Learning
RMSE	Root Mean Square Error
R^2	Coefficient of Determination
SDG	Sustainable Development Goal
UI	User Interface
XGBoost	Extreme Gradient Boosting

LIST OF APPENDICES

Appendix A: Survey instruments	36
Appendix B: Risk analysis.....	36
Appendix C: Ethical documents	37

1. INTRODUCTION

1.1 Background and Context

Municipal solid waste management has become a critical environmental and operational challenge in rapidly urbanizing regions around the world. Increasing population density, economic development, and changing consumption patterns have significantly increased the volume and diversity of waste generated in cities. Developing countries in particular face considerable difficulties in managing waste effectively due to limited infrastructure, inefficient operational planning, and insufficient adoption of data-driven technologies [1]. In many regions, waste collection systems still rely on fixed schedules and reactive operational strategies rather than intelligent prediction and planning mechanisms.

Sri Lanka faces similar challenges in municipal waste management. According to national environmental reports, the country generates approximately 7,000–8,000 metric tons of municipal solid waste per day, with a significant portion originating from urban regions such as Colombo and other metropolitan areas [2]. A large share of this waste consists of organic materials, plastics, paper, and other recyclable resources. However, effective waste segregation and recycling remain limited due to inadequate planning and insufficient data-driven decision support within waste management operations. Recycling facilities and municipal authorities often lack advance knowledge regarding the composition of waste that will arrive during upcoming collection cycles. As a result, processing facilities may become overloaded with certain waste types while other resources remain underutilized.

These inefficiencies affect multiple stakeholders including municipal councils, waste management service providers, recycling facilities, and urban communities. Poorly planned waste collection operations increased fuel consumption, operational costs, and landfill dependency while reducing recycling efficiency and environmental sustainability [3]. Furthermore, unmanaged waste accumulation may contribute to environmental pollution, health risks, and deterioration of urban living conditions.

Recent advancements in smart waste management have introduced Internet of Things (IoT) technologies that enable real-time monitoring of waste bins using sensors to measure fill level, weight, temperature, and gas emissions [4]. These systems improve operational visibility by allowing waste management authorities to detect overflow conditions and optimize collection routes. However, most existing solutions remain reactive, as they primarily monitor waste conditions only after waste has been disposed. While research has explored route optimization and bin monitoring, limited attention has been given to forecasting the composition of waste before collection occurs.

To address this limitation, this research proposes a Machine Learning-based Waste Composition Forecasting System that predicts the expected distribution of waste categories for the next collection cycle. The proposed approach utilizes historical waste collection records together with contextual variables such as weather conditions, calendar information, and location characteristics to identify patterns influencing waste generation. Machine learning algorithms such as Random Forest and XGBoost can analyze these patterns and

forecast the percentage composition of waste types including organic waste, plastic, paper, and metal. By providing predictive insights before collection takes place, waste management organizations can better prepare recycling operations and allocate resources efficiently.

This research also aligns with global sustainability initiatives, particularly the United Nations Sustainable Development Goals (SDGs). The proposed system supports SDG 11: Sustainable Cities and Communities by improving urban waste management efficiency and SDG 12: Responsible Consumption and Production by promoting better recycling practices and resource utilization [5]. From a research perspective, this study falls within the Technology Integration and Management cluster in Information Technology, where predictive analytics techniques are applied to environmental and sustainability challenges.

Overall, the proposed forecasting system aims to transform waste management from a reactive process into a proactive and data-driven operation. By predicting waste composition before collection, the system has the potential to enhance recycling preparation, improve operational efficiency, and contribute to sustainable urban waste management practices.

1.2 Literature Review

Recent literature shows that smart waste management research has gradually moved from conventional waste collection models toward data-driven and predictive systems. Early forecasting studies mainly concentrated on estimating total municipal solid waste generation rather than the composition of waste streams. Abbasi and El Hanandeh developed artificial intelligence-based models for municipal solid waste forecasting and showed that machine learning methods can outperform traditional forecasting approaches in handling nonlinear waste generation patterns [6]. Similarly, Kannangara *et al.* demonstrated that machine learning techniques can effectively predict regional municipal solid waste generation and diversion using demographic and socioeconomic variables, highlighting the value of data-driven prediction in planning waste operations [7].

More recent studies have expanded the forecasting problem by incorporating broader contextual and spatial factors. Lu *et al.* proposed a multi-city machine learning model for municipal solid waste generation prediction, showing that predictive performance can be improved when models are trained using data from multiple urban contexts rather than a single locality [8]. Wan *et al.* further showed that district-level prediction can provide more useful decision support for urban authorities by capturing local variations in waste generation patterns [9]. These studies confirm that machine learning is increasingly recognized as a practical tool for waste forecasting, especially where waste behavior varies across regions and time periods.

A second line of research has investigated the role of advanced learning models such as artificial neural networks and deep learning architectures. Puntarić *et al.* used an artificial neural network model to predict separated municipal solid waste fractions and found that ANN-based approaches can produce useful estimates when socio-demographic and economic factors are available [10]. Zhang *et al.* proposed a multi-factor deep learning model based on BiLSTM for municipal solid waste generation forecasting and reported that temporal models can better capture dynamic relationships among influencing variables [11]. Likewise, Liu *et al.* applied an LSTM-based model using multiple input indicators and showed that deep learning can improve forecasting accuracy in complex urban waste systems [12]. These

findings are important because they suggest that temporal and multi-factor models may be suitable for predicting future waste conditions when historical and contextual data are available.

In parallel with forecasting research, a substantial body of literature has focused on smart waste management systems based on IoT and sensor-enabled monitoring. Cheema *et al.* reviewed smart waste management and classification systems and found that most current approaches emphasize bin-level sensing, route optimization, and waste segregation using machine learning or deep learning techniques [13]. Fang *et al.* also reported that artificial intelligence in smart-city waste management is being applied extensively in waste-to-energy systems, smart bins, sorting robotics, monitoring platforms, and waste generation models, but the maturity of these solutions varies across application domains [14]. These studies indicate that although predictive analytics is gaining attention, real-world implementations are still largely centered on monitoring waste after disposal rather than forecasting waste composition before collection.

Recent review studies reinforce this observation. Pillai and Singh noted that machine learning in municipal solid waste management has mostly been used for automation of collection, transportation, classification, and generation forecasting, while integrated predictive systems remain limited [15]. Dawar *et al.* similarly concluded that machine learning and deep learning are increasingly used for waste generation forecasting, route planning, and recycling support, but integration across the waste management lifecycle is still incomplete [16]. This is particularly relevant to the present study, because it suggests that the research field has not yet sufficiently addressed predictive models that estimate next-day waste composition specifically for operational recycling preparation.

Existing products and services in the market also reveal a similar pattern. Commercial solutions such as Bigbelly, Sensoneo, Ecube Labs, and Compology provide smart waste platforms with features such as real-time fill-level monitoring, compaction, bin analytics, route planning, and container visibility [17] – [20]. These systems have clear operational value because they reduce overflow, optimize collection frequency, and improve monitoring of distributed waste assets. However, their primary emphasis remains on sensing current bin status and optimizing collection logistics. They do not explicitly address the problem of forecasting the expected composition of waste categories for the next collection cycle. As a result, their support for proactive recycling preparation is limited.

Overall, the literature demonstrates strong progress in three areas: municipal waste forecasting, intelligent waste monitoring, and smart operational platforms. However, most studies either predict total waste quantity, classify waste after disposal, or optimize collection after waste accumulation has already occurred [6] - [16]. Likewise, commercial systems mainly support monitoring and logistics rather than pre-collection forecasting of waste composition [17] – [20]. Therefore, there remains a clear need for a machine learning-based approach that predicts the next-day composition of waste categories using historical and contextual data, thereby enabling proactive recycling planning and resource allocation.

1.3 Research Gap

Although significant progress has been made in the development of smart waste management systems, several limitations remain in current research and practical implementations. Existing studies largely concentrate on improving waste monitoring, collection logistics, and waste quantity forecasting rather than predicting the composition of waste prior to collection. For example, many machine learning-based studies focus on forecasting the total amount of municipal solid waste generated in urban regions using demographic, socioeconomic, or environmental variables [6] – [9]. While such models are useful for long-term planning and policy development, they typically estimate overall waste generation volumes rather than the distribution of waste categories, which is essential for operational recycling preparation.

Furthermore, several recent studies have explored advanced prediction techniques such as artificial neural networks and deep learning architectures for municipal waste forecasting [10] – [12]. These models demonstrate improved prediction accuracy by capturing complex relationships between influencing variables and waste generation patterns. However, their primary focus remains on predicting waste quantity or general trends rather than identifying the expected composition of waste types such as organic material, plastics, paper, and metals. Consequently, these approaches provide limited operational value for recycling facilities that require advance knowledge of material composition in order to prepare sorting and processing activities.

Another major limitation is evident in current smart waste management systems deployed in practice. Most commercial solutions, including platforms such as Bigbelly, Sensoneo, Ecube Labs, and Compology, primarily rely on IoT sensors to monitor bin fill levels, detect overflow conditions, and optimize collection routes [17] – [20]. While these technologies provide valuable real-time operational visibility, they remain fundamentally reactive systems because they respond only after waste has already been deposited into bins. As a result, recycling facilities and waste processing centers still lack predictive insights regarding the composition of waste that will arrive during the next collection cycle.

The absence of predictive models for next-day waste composition forecasting represents a significant research gap in the field of intelligent waste management. Accurate prediction of waste composition would allow waste management organizations to prepare recycling equipment, allocate human resources, and schedule processing activities before waste arrives at processing facilities. Without such predictive capabilities, recycling operations must rely on assumptions or historical averages, which may lead to inefficient resource allocation and reduced recycling efficiency.

Therefore, this research proposes a Machine Learning-based Waste Composition Forecasting System designed to predict the expected distribution of waste categories for upcoming collection cycles. By combining historical waste data with contextual variables such as weather conditions, calendar information, and location characteristics, the proposed model aims to generate reliable predictions of waste composition. This approach introduces a proactive decision-support mechanism that complements existing monitoring systems and extends current waste management research beyond reactive monitoring toward predictive operational planning.

The novelty of this research lies in its focus on predicting waste composition before collection occurs, enabling recycling facilities to prepare appropriate processing strategies in

advance. By addressing the identified limitations in existing forecasting models and smart waste monitoring platforms, the proposed system contributes a new predictive intelligence layer to smart waste management frameworks and supports more efficient, sustainable, and data-driven waste management practices.

Table 1: Research Gap Summary table

Research Area	Existing Approaches	Identified Limitations	Proposed Contribution
Waste Generation Forecasting	Many studies use machine learning models to forecast the total volume of municipal solid waste using demographic, socioeconomic, and environmental variables [6] – [9].	These models mainly estimate overall waste quantity and provide limited insight into the distribution of waste types required for recycling operations.	The proposed research utilizes waste volume data together with contextual variables to estimate and predict waste composition patterns.
Deep Learning-Based Waste Prediction	Advanced techniques such as artificial neural networks and deep learning models are applied to improve prediction accuracy of waste generation trends [10] – [12].	Most models focus on predicting waste quantity rather than the material composition of waste streams.	This research extends prediction capabilities by integrating contextual data to forecast the likely composition of waste categories.
Smart Waste Monitoring Systems	Commercial systems such as Bigbelly, Sensoneo, Ecube Labs, and Compology provide IoT-based monitoring of bin fill levels and collection logistics [17]– [20].	These systems provide real-time monitoring but operate reactively after waste is deposited in bins.	The proposed system introduces predictive analytics to support planning before waste collection occurs.
Recycling Operation Planning	Waste management decisions are often based on historical averages or manual estimation methods.	Lack of predictive insights about waste composition can lead to inefficient resource allocation and recycling preparation.	The proposed system provides data-driven predictions to assist recycling facilities in preparing appropriate processing strategies.
Intelligent Waste Management Research	Most intelligent waste management studies focus on IoT monitoring, route optimization, and waste quantity estimation.	Limited research integrates waste volume data, contextual variables, and machine learning to support predictive operational planning.	The proposed research integrates waste volume records, contextual factors, and machine learning models to provide improved predictive decision support.

2. OBJECTIVE

2.1 Main Objective

The main objective of this research is to design, develop, and evaluate a machine learning–based waste composition forecasting system capable of predicting the next-day distribution of major waste categories (organic, plastic, paper, and metal) using historical waste collection data and contextual factors such as weather conditions, calendar information, and location characteristics.

The proposed system aims to generate accurate waste composition predictions that can support proactive recycling preparation and resource allocation in waste management operations. The performance of the forecasting model will be evaluated using established machine learning evaluation metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to assess prediction accuracy and model reliability.

2.2 Specific Objectives

- To collect and preprocess historical waste collection data and relevant contextual datasets, including weather information, calendar events, and location characteristics, in order to create a structured dataset suitable for machine learning analysis.
- To identify and construct meaningful predictive features such as area type, date type (weekday, weekend, and holiday), weather conditions, and historical waste generation patterns that influence waste composition.
- To design and train machine learning models, such as Random Forest and XGBoost, capable of predicting the percentage composition of waste categories including organic waste, plastic, paper, and metal.
- To evaluate the performance and reliability of the developed prediction models using appropriate machine learning evaluation metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and cross-validation techniques.
- To develop a visualization interface or dashboard that presents the predicted waste composition results to support decision-making for recycling preparation and resource allocation.
- To analyze how predictive waste composition insights can improve operational planning, particularly in preparing recycling processes and optimizing resource utilization in waste management systems.

3. METHODOLOGY

This research adopts a data-driven approach to design and evaluate a machine learning-based system capable of predicting next-day waste composition. The methodology combines data collection, preprocessing, model development, and experimental validation to generate reliable predictive insights that can support proactive recycling preparation in waste management systems.

3.1 Research Approach

The study follows a Design Science Research (DSR) approach combined with experimental machine learning evaluation. Design Science Research focuses on developing and evaluating technological artifacts that address real-world problems. In this research, the artifact is a predictive waste composition forecasting model that supports decision-making in waste management operations.

The research process consists of several phases: problem identification, data preparation, predictive model development, experimental evaluation, and performance validation. Machine learning techniques are applied to identify patterns within historical waste data and contextual variables. Experimental evaluation is then conducted to measure the prediction accuracy and reliability of the developed models.

3.2 Technical Methods and Algorithms

The proposed system utilizes supervised machine learning techniques to predict the percentage distribution of different waste categories. The prediction task is treated as a regression problem where the model estimates the proportion of waste categories such as organic waste, plastic, paper, and metal.

Two machine learning algorithms are selected for this study:

Random Forest Regression:

Random Forest is an ensemble learning algorithm that constructs multiple decision trees and aggregates their predictions to improve accuracy and reduce overfitting. It is widely used in environmental prediction tasks due to its ability to handle nonlinear relationships and heterogeneous datasets [7].

Extreme Gradient Boosting (XGBoost):

XGBoost is a gradient boosting algorithm that improves predictive performance by iteratively optimizing decision trees. It has been successfully applied in predictive analytics and environmental forecasting because of its high accuracy and computational efficiency [6].

These algorithms will analyze historical waste composition data together with contextual factors such as weather conditions, date information, and location characteristics to identify patterns influencing waste generation.

3.3 Tools and Platforms

The implementation of the proposed system will utilize widely adopted data science tools and machine learning frameworks.

Table 2: Technologies and Tools that will be used

System Component	Tools / Technologies
Programming Language	Python
Data Processing Libraries	Pandas, NumPy
Machine Learning Frameworks	Scikit-learn, XGBoost, TensorFlow
Data Visualization	Matplotlib, Seaborn, Plotly
Dashboard / Web Interface	Streamlit, MERN Stack (MongoDB, Express, React, Node.js), Flask
Data Storage	MongoDB
External Data Sources	Weather APIs (OpenWeatherMap, WeatherStack)
Development Environment	Jupyter Notebook, Google Colab, VS Code
Deployment Platform	Cloud platforms such as AWS, Microsoft Azure, or Google Cloud

These technologies are selected due to their reliability, strong community support, and suitability for rapid development of machine learning-based systems.

3.4 Data Collection and Ethical Considerations

The predictive model requires multiple sources of data to identify patterns in waste generation. The dataset will include:

- Historical waste composition records obtained from waste collection logs or publicly available datasets.
- Contextual environmental data including temperature, rainfall, and humidity obtained from weather APIs.
- Calendar information such as weekdays, weekends, public holidays, and special events.
- Location characteristics including area type (e.g., residential, campus, or commercial zones).

Where real-world historical data is limited, simulated datasets may be generated using realistic waste statistics and distributions reported in previous research studies [8].

From an ethical perspective, the research primarily uses environmental and operational data rather than personal or sensitive information. Therefore, the project presents minimal privacy risks. However, appropriate data management practices will be followed to ensure secure storage, responsible usage, and compliance with institutional research guidelines.

3.5 Validation Strategy and Performance Metrics

The predictive models will be evaluated using standard machine learning validation techniques to assess accuracy and reliability.

Cross-Validation:

K-fold cross-validation will be used to ensure that the model generalizes well to unseen data and to prevent overfitting.

Performance Metrics:

The prediction accuracy of the models will be measured using the following metrics:

- **Mean Absolute Error (MAE):** measures the average absolute difference between predicted and actual values.
- **Root Mean Square Error (RMSE):** evaluates prediction accuracy by penalizing larger errors.
- **Coefficient of Determination (R^2):** measures how well the model explains the variance in the dataset.

Comparative analysis will be conducted between Random Forest and XGBoost models to determine which algorithm provides the best performance for waste composition prediction.

Through this methodology, the research aims to develop a reliable machine learning forecasting system capable of supporting proactive recycling planning and improving operational efficiency in waste management systems.

4. HIGH-LEVEL SYSTEM ARCHITECTURE

4.1 Overview of the Proposed System

The proposed Waste Composition Forecasting System is designed as a data-driven predictive analytics module within a smart waste management framework. The primary purpose of this component is to forecast the expected distribution of waste categories for the next collection cycle using historical waste data and contextual information. By predicting the composition of waste types in advance, waste management organizations can prepare recycling processes, allocate resources efficiently, and improve operational planning.

The architecture of the proposed system follows a layered structure consisting of five major components: data sources, data processing module, machine learning prediction engine, forecasting output module, and visualization interface. These components collectively enable the system to collect relevant data, process and analyze it using machine learning techniques, generate predictive insights, and present the results in a user-friendly format for decision support.

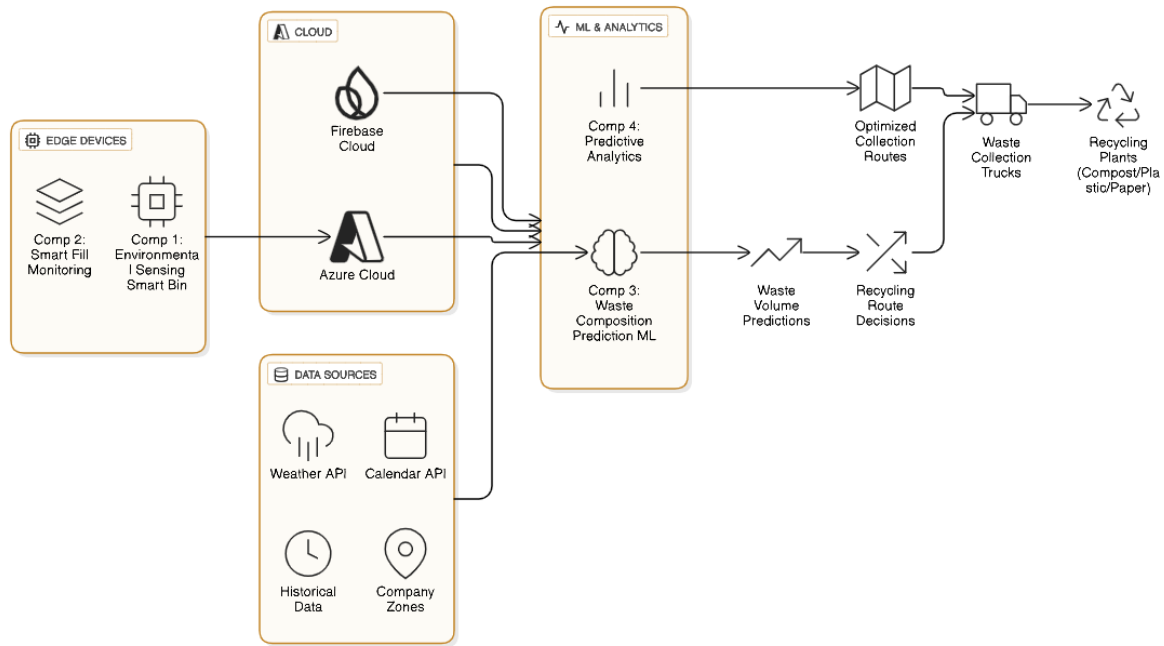


Figure 1: High-Level Architecture Diagram of the System

4.2 System Components

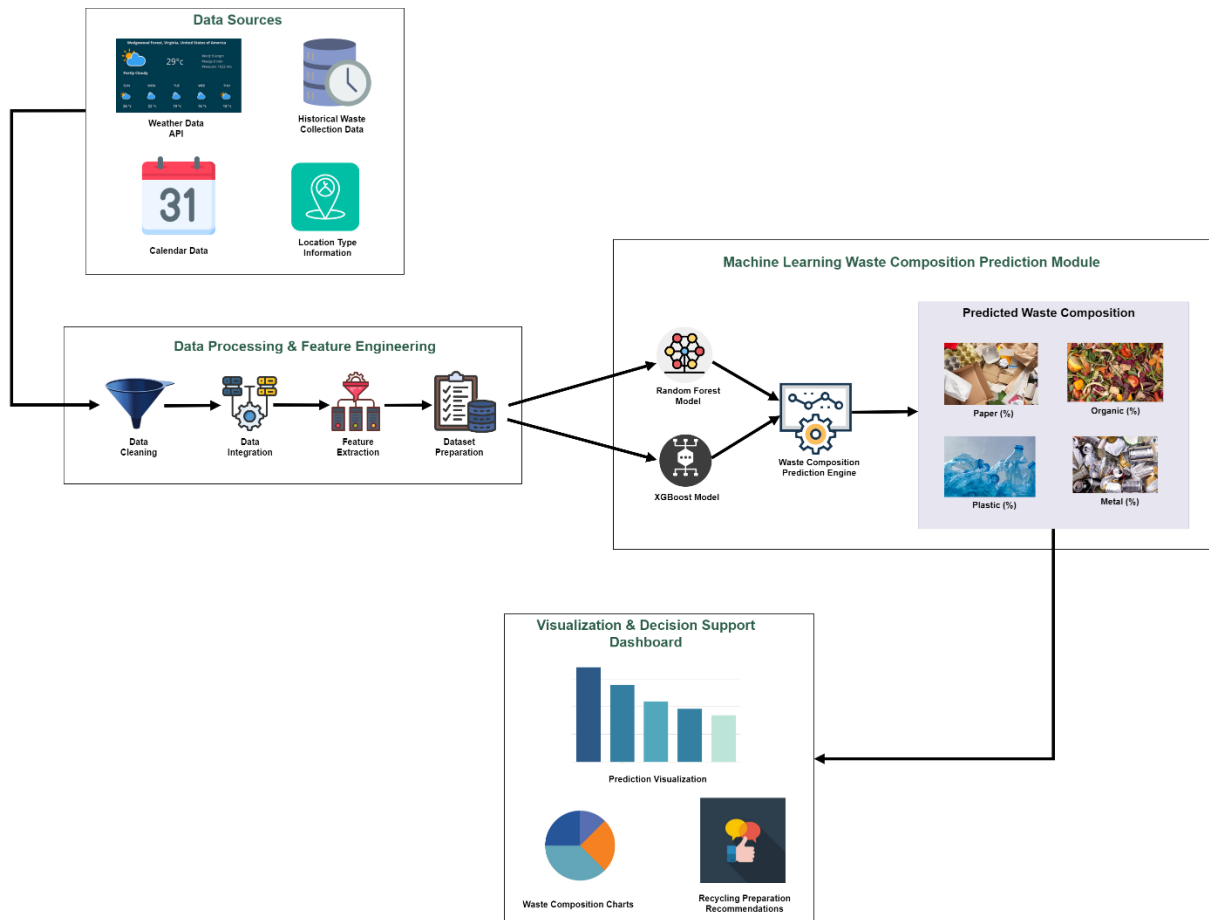


Figure 2: Component Architecture Diagram

The architecture of the proposed forecasting system consists of the following main components:

Data Sources Layer:

This layer contains multiple sources of input data used for prediction. The system collects historical waste collection records, contextual weather data, calendar information, and location characteristics. These datasets provide the contextual factors necessary for identifying patterns in waste generation.

Data Processing and Feature Engineering Module:

This component performs preprocessing tasks such as data cleaning, normalization, and transformation. Feature engineering techniques are applied to convert raw data into meaningful predictive features, such as weekday indicators, weather attributes, and historical waste composition patterns.

Machine Learning Prediction Engine:

The prediction engine is the core component of the system. It utilizes supervised machine learning algorithms such as Random Forest and XGBoost to analyze historical data and generate predictions for the expected distribution of waste categories. The trained models learn relationships between contextual variables and waste composition patterns.

Forecasting Output Module:

Once the prediction is generated, the system produces the estimated percentage distribution of different waste types such as organic waste, plastic, paper, and metal. These predictions are stored in the system and used for operational planning and decision support.

Visualization and Decision Support Dashboard:

A web-based dashboard presents the prediction results in a clear and interpretable format. The dashboard allows waste management authorities to view predicted waste composition, identify trends, and prepare recycling operations accordingly.

4.3 Data Flow Explanation

The data flow within the proposed system follows several sequential stages:

1. Raw data is collected from historical waste records, weather APIs, and calendar data sources.
2. The collected data is stored and preprocessed to remove inconsistencies and create structured datasets.
3. Feature engineering techniques generate predictive variables that represent contextual conditions influencing waste generation.
4. The machine learning model processes the dataset and produces predictions for next-day waste composition.
5. The predicted results are transmitted to the visualization interface where they are displayed to users for operational planning.

This data flow ensures that information moves efficiently from raw input sources to decision-support outputs.

4.5 Scalability and Security Considerations

The proposed architecture is designed to be scalable and adaptable for larger waste management environments. Since the predictive system relies on cloud-compatible data processing tools and machine learning frameworks, the architecture can scale to support larger datasets, multiple geographical zones, and continuous data updates. Cloud storage and processing platforms can be integrated to handle increasing volumes of waste data without compromising system performance.

Security and data integrity are also important considerations. Although the system primarily processes environmental and operational data rather than personal information, appropriate data protection measures must be implemented. Secure APIs should be used for external data access, and access control mechanisms should be implemented within the system dashboard to prevent unauthorized modifications. Additionally, regular data backups and monitoring mechanisms can ensure reliability and prevent data loss.

Overall, the proposed architecture provides a scalable and secure framework capable of supporting predictive waste composition analysis within modern smart waste management systems.

5. USER REQUIREMENTS

5.1 Stakeholder Identification

The proposed waste composition forecasting system will support several stakeholders involved in waste management operations and recycling processes. These stakeholders interact with the system to obtain predictive insights that improve operational planning and resource utilization.

Municipal Waste Management Authorities

Municipal councils and government waste management departments are responsible for organizing waste collection, recycling programs, and landfill management. These authorities can use the system to analyze predicted waste composition and prepare appropriate recycling strategies.

Waste Collection Service Providers

Public or private waste collection companies manage waste transportation and collection schedules. Access to predicted waste composition data can assist these organizations in planning collection logistics and coordinating with recycling facilities.

Recycling Facilities and Waste Processing Centers

Recycling plants and material recovery facilities require advance knowledge about the type and quantity of incoming waste streams. The forecasting system allows these facilities to prepare appropriate sorting processes, equipment, and workforce allocation.

Smart City and Urban Planning Authorities

Urban planners and smart city initiatives can use predictive waste analytics to develop more sustainable waste management strategies and improve long-term infrastructure planning.

5.2 Functional Requirements

FR1: The system shall collect and integrate historical waste composition data from waste management records.

FR2: The system shall retrieve contextual environmental data such as weather conditions using external weather APIs.

FR3: The system shall incorporate calendar-based contextual information including weekdays, weekends, holidays, and special events.

FR4: The system shall preprocess and transform collected data into structured datasets suitable for machine learning analysis.

FR5: The system shall train machine learning models such as Random Forest and XGBoost using historical datasets.

FR6: The system shall generate predictions for next-day waste composition categories including organic waste, plastic, paper, and metal.

FR7: The system shall present predicted waste composition results through an interactive visualization dashboard.

FR8: The system shall allow users to view predicted waste composition percentages and related analytical insights to support recycling preparation.

5.3 Non-Functional Requirements

Security Requirements

The system shall ensure secure access to the prediction dashboard using authentication mechanisms. Data obtained from external APIs should be transmitted using secure communication protocols such as HTTPS to prevent unauthorized access.

Performance Requirements

The system shall generate prediction results within an acceptable time frame to support operational planning. Machine learning models should be optimized to ensure efficient processing of datasets without excessive computational delays.

Scalability Requirements

The architecture shall support expansion to handle larger datasets and multiple geographic regions as the system grows. Cloud-based infrastructure may be integrated to support increased data storage and computational demands.

Usability Requirements

The dashboard interface shall present prediction results in a clear and understandable format using charts and visual indicators. Users without advanced technical knowledge should be able to interpret prediction outputs easily.

Reliability Requirements

The system shall maintain consistent prediction performance and ensure reliable data processing. Data backup and system monitoring mechanisms should be implemented to prevent data loss and maintain system stability.

6. COMMERCIALIZATION PLAN

6.1 Target Market and Customer Persona

The proposed Waste Composition Forecasting System has potential commercial applications for organizations responsible for waste collection, recycling, and sustainable environmental management. The system is particularly valuable for institutions that require better planning and optimization of waste processing operations.

Table 3: Target Market Segments

Target Market	Description	Potential Benefits
Municipal Councils	Local government authorities responsible for city waste management	Improved recycling planning and resource allocation
Private Waste Management Companies	Organizations providing waste collection and transportation services	Operational efficiency and cost reduction
Recycling Facilities	Material recovery and recycling plants	Better preparation for incoming waste streams
Universities & Campuses	Large institutions managing internal waste systems	Data-driven sustainability initiatives

Smart City Projects	Urban infrastructure projects using digital technologies	Integration with smart city environmental systems
---------------------	--	---

Table 4: Customer Persona

Persona	Role	Responsibilities	System Benefits
Waste Management Operations Manager	Municipal authority or waste company	Oversees waste collection and recycling operations	Predicts waste composition for operational planning
Environmental Sustainability Officer	Large organizations or institutions	Manages sustainability and waste reduction programs	Uses predictive analytics for recycling improvement
Recycling Facilities	Recycling plants	Prepares sorting and processing operations	Plans equipment and workforce allocation

6.2 Value Proposition

The proposed system provides a data-driven predictive solution that enables proactive waste management planning. Unlike traditional systems that only monitor waste levels after disposal, this system predicts the composition of waste expected in the next collection cycle.

Key value propositions include:

- Improved recycling preparation through advance knowledge of waste composition
- Reduced operational costs by optimizing resource allocation
- Increased recycling efficiency and reduced landfill dependency
- Data-driven decision support for waste management authorities
- Contribution to sustainable and environmentally responsible waste management

These benefits provide both operational and environmental value to organizations responsible for managing waste streams.

6.3 Revenue Model

The system can be commercialized using a Software-as-a-Service (SaaS) business model where customers subscribe to the forecasting platform.

Table 5: Potential Revenue Streams

Revenue Source	Description
Subscription Service	Monthly or annual subscription for platform access
Data Analytics Services	Customized waste analytics for organizations
System Integration	Integration with municipal or smart city waste systems
Consulting Services	Environmental planning and waste management optimization

6.4 Cost Estimation

The development and operational costs of the proposed system mainly involve software development, data infrastructure, and system maintenance. Key cost components include:

- Cloud storage and data processing infrastructure
- API subscription costs for weather and environmental data
- Software development and system integration
- Maintenance and technical support

Since the system relies primarily on software and publicly available data sources, the overall operational costs are relatively low compared to hardware-intensive waste monitoring systems.

6.5 Pricing Strategy

A tiered subscription pricing model can be used to accommodate different customer segments.

For example:

- **Basic Plan:** Small institutions or campuses with limited waste monitoring needs
- **Standard Plan:** Municipal councils or medium-sized waste management organizations
- **Enterprise Plan:** Large cities or integrated smart waste management systems requiring advanced analytics and multiple-zone predictions

Pricing can vary depending on the scale of deployment, number of monitored locations, and level of predictive analytics provided.

6.6 Competitive Advantage

Most existing smart waste management systems focus on real-time monitoring and route optimization. However, they rarely provide predictive insights about waste composition before collection occurs.

Table 6: Comparison with Existing Solutions

Feature	Existing Smart Waste Systems	Proposed System
Real-Time Bin Monitoring	✓	✓
Route Optimization	✓	✓
Waste Composition Prediction	✗	✓
Pre-Collection Recycling Planning	✗	✓
Data-Driven Forecasting	Limited	Advanced ML-based forecasting

The predictive forecasting capability provides a unique advantage by enabling proactive waste management planning.

6.7 Intellectual Property Considerations

The proposed forecasting system may involve intellectual property related to the predictive modeling approach, data processing methods, and software architecture. The developed algorithms and software implementation could potentially be protected through software copyright or intellectual property registration.

If the system demonstrates strong commercial potential, organizations may consider applying for software patents or proprietary analytics models to protect the predictive framework. Additionally, the system architecture and data analytics methodology may be treated as proprietary technology within a commercial platform.

Proper documentation and licensing strategies can ensure that the intellectual property associated with the forecasting system is protected while allowing commercialization opportunities.

7. BUDGET AND JUSTIFICATION

The development of the proposed Machine Learning–Based Waste Composition Forecasting System requires several resources including computing hardware, software tools, cloud infrastructure, and development effort. Since the proposed system is mainly software-oriented and uses open-source technologies, the overall implementation cost remains relatively affordable.

Hardware Costs

The system requires a development computer for machine learning model training, data processing, and dashboard implementation. Additional storage and internet connectivity may also be required for data access and backups.

Table 7: Estimated Hardware Costs

Hardware Resource	Description	Estimated Cost (Rs)
Existing Laptop	Used for programming, machine learning model training, and data analysis	Already available
RAM Upgrade (Optional)	Additional memory to support machine learning model execution	15,000
External Storage	Backup storage for datasets and project files	10,000
Estimated Hardware Upgrade Cost: Rs.25,000		

A reliable computing device is required to run machine learning algorithms and perform data analysis. External storage ensures safe backup of datasets and project resources. Internet connectivity is essential for accessing weather APIs and cloud services used in the system.

Software Licenses

The system will primarily rely on open-source software tools, which significantly reduces development costs.

Table 8: Software Licenses Costs

Software Tool	Purpose	License Cost
Python	Programming language for system development	Free
Pandas / NumPy	Data preprocessing and numerical computation	Free
Scikit-learn	Machine learning model implementation	Free
XGBoost	Gradient boosting algorithm for prediction	Free
Streamlit	Web dashboard development	Free
Diagram Tools (Draw.io / Lucidchart)	System architecture diagrams	Free / Optional
Total Software Cost: Rs.0		

The selected technologies are widely used open-source tools in machine learning and data science research. Their availability significantly reduces project development costs while still providing powerful functionality for predictive analytics.

Cloud Services

Cloud infrastructure may be used to store datasets, process prediction models, and host the visualization dashboard.

Table 9: Cloud Services Costs

Cloud Service	Purpose	Estimated Monthly Cost (Rs)
Cloud Storage (Firebase / AWS / Azure)	Dataset storage and management	3,000
Cloud Computing Services	Running machine learning predictions	4,000
Weather API Services	Access to weather data	Free / minimal
Estimated Monthly Cloud Cost: Rs.7,000		

Cloud services provide scalable computing and storage resources, which are useful for handling larger datasets and hosting the forecasting dashboard. However, many cloud providers offer free-tier services that can support prototype research systems.

Human Effort Estimation

Table 10: Estimated Human Effort hours

Development Activity	Estimated Hours
Data collection and preprocessing	130 hours
Feature engineering and dataset preparation	72 hours
Machine learning model development	246 hours
Model evaluation and validation	72 hours
Dashboard development and visualization	80 hours
Total Estimated Development Effort: 600 hours	

The estimated hours represent the time required to design, develop, and evaluate the forecasting system. This includes data preparation, machine learning model development, system implementation.

8. WORK BREAKDOWN STRUCTURE (WBS)

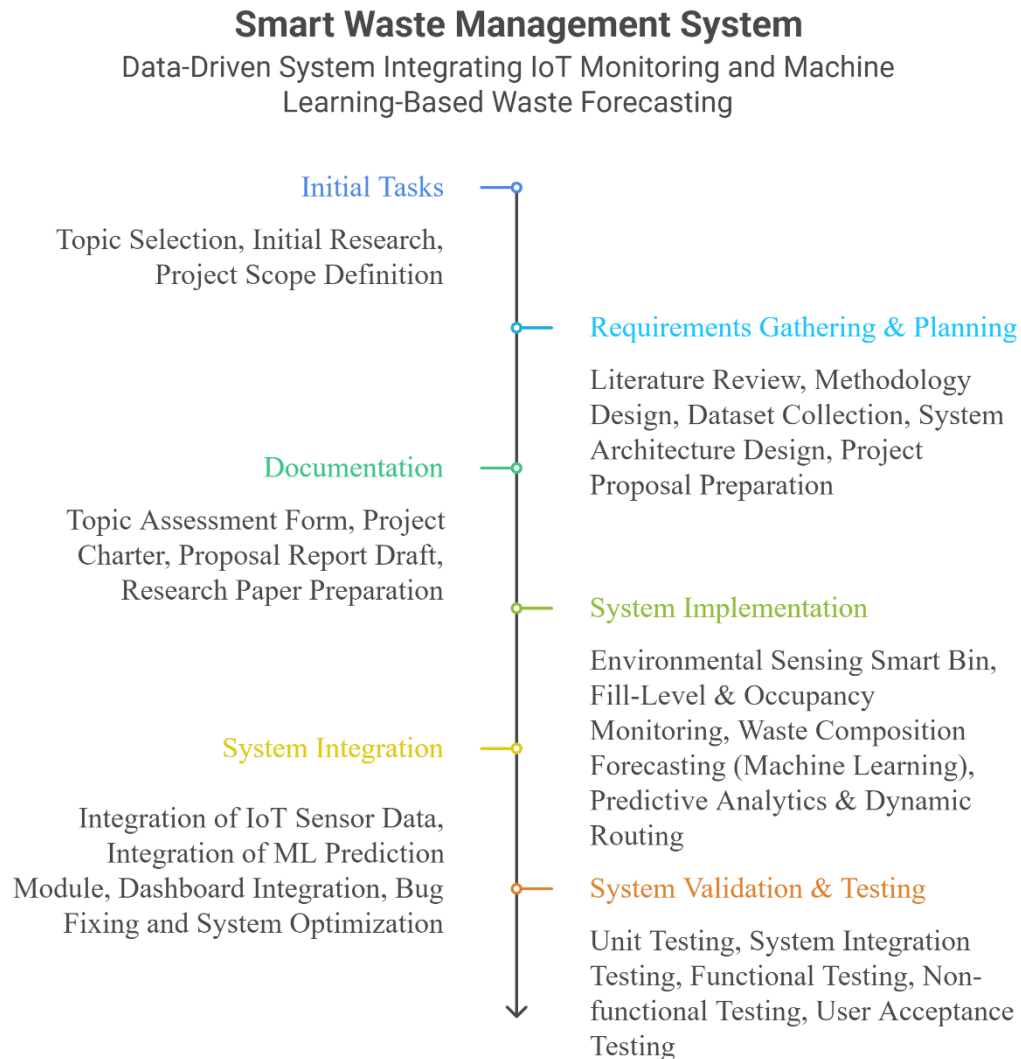


Figure 3: Work Breakdown Structure of the overall project

The Work Breakdown Structure (WBS) outlines the distribution of tasks among the project team members for the development of the Smart Waste Management System. Each member is responsible for a specific component of the system, ensuring balanced workload distribution and efficient project execution.

Table 11: Estimated Human Effort hours

Task ID	Task Description	Timeline	Deliverable
T1	Collection of historical waste data and contextual datasets (weather, calendar, location)	March – April	Initial dataset
T2	Data preprocessing and cleaning	April	Cleaned dataset
T3	Feature engineering and dataset preparation	April – May	Machine learning dataset
T4	Development of machine learning prediction models (Random Forest / XGBoost)	May – June	Initial ML models
T5	Model training and hyperparameter tuning	June – July	Optimized prediction model
T6	Model evaluation using MAE, RMSE, and cross-validation	July – August	Model evaluation results
T7	Development of prediction visualization dashboard	August – September	Dashboard prototype
T8	System testing and performance analysis	September – October	System testing results

Table 12: Workload Distribution

Development Phase	Tasks	Estimated Effort
Data Preparation	T1 – T3	40%
Model Development	T4 – T5	30%
Evaluation & Testing	T6 – T8	20%

9. GANTT CHART

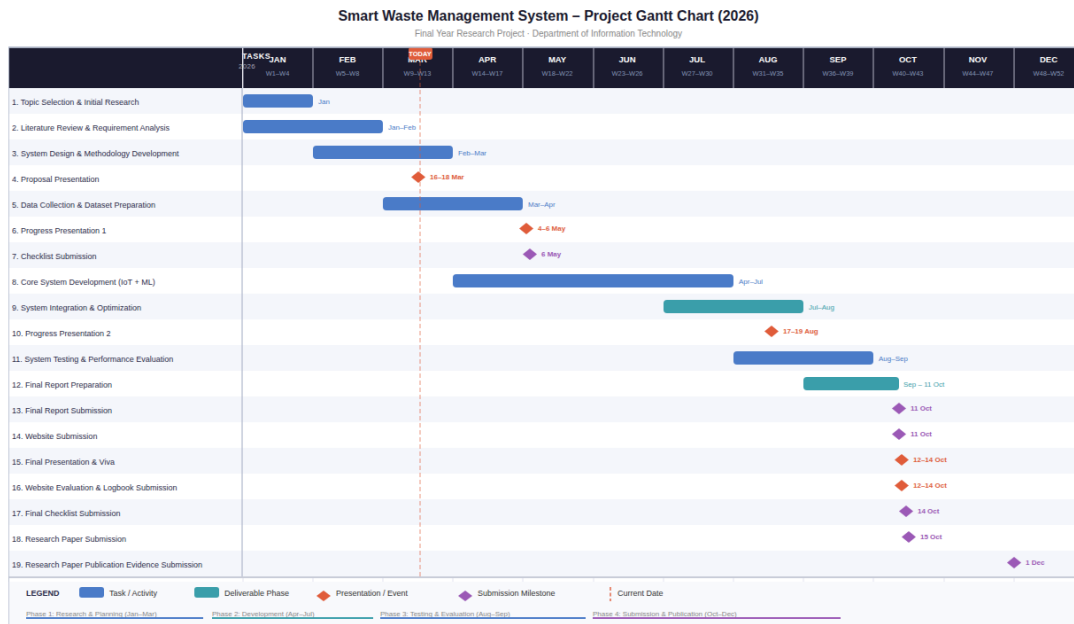


Figure 4: Gantt chart

10. REFERENCES LIST

- [1] K. 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] Ministry of Environment Sri Lanka, *National Waste Management Policy and Strategy of Sri Lanka*. Colombo, Sri Lanka, 2020.
- [3] United Nations Environment Programme (UNEP), *Global Waste Management Outlook*. Nairobi, Kenya, 2015.
- [4] M. S. Kumar and P. Dutta, "Smart city waste management using sensor-based IoT architecture," *Journal of Wireless Sensor Networks and IoT*, vol. 6, no. 2, pp. 45-56, 2020.
- [5] United Nations, *Transforming our world: The 2030 Agenda for Sustainable Development*. New York, NY, USA, 2015.
- [6] M. Abbasi and A. El Hanandeh, "Forecasting municipal solid waste generation using artificial intelligence modelling approaches," *Waste Management*, vol. 56, pp. 13–22, Oct. 2016.
- [7] M. Kannangara, R. Dua, L. Ahmadi, and F. Bensebaa, "Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches," *Waste Management*, vol. 74, pp. 3–15, Apr. 2018.
- [8] W. Lu, Y. Chen, C. Webster, X. Wang, and Z. Zhang, "Development of machine learning multi-city model for municipal solid waste generation prediction," *Frontiers of Environmental Science & Engineering*, vol. 16, 2022.
- [9] S. Wan, Y. Li, and X. Chen, "Insights into the urban municipal solid waste generation forecasting: A district-level machine learning approach for New York City," *Sustainable Cities and Society*, vol. 102, 2024.
- [10] E. Puntarić, N. Vukojević Medvidović, and I. Prusina, "Prediction of the production of separated municipal solid waste using artificial neural networks," *Sustainability*, vol. 14, no. 16, Art. no. 10133, 2022.
- [11] X. Zhang, Y. Li, and H. Zhao, "Prediction and feed-in tariffs of municipal solid waste power generation based on deep learning," *Sustainability*, vol. 16, no. 9, Art. no. 3579, 2024.
- [12] B. Liu, Y. Zhang, and J. Wang, "Prediction of municipal solid waste generation and associated environmental factors using LSTM-based modelling," *Environmental Earth Sciences*, vol. 83, 2024.
- [13] S. M. Cheema, A. Hannan, and I. M. Pires, "Smart waste management and classification systems using cutting edge approach," *Sustainability*, vol. 14, no. 16, Art. no. 10226, 2022.
- [14] B. Fang, Y. Li, and X. Zhou, "Artificial intelligence for waste management in smart cities: A review," *Environmental Chemistry Letters*, vol. 21, 2023.

- [15] K. S. Pillai and R. Singh, "Municipal solid waste management: A review of machine learning applications," *E3S Web of Conferences*, vol. 447, 2023.
- [16] I. Dawar, M. A. Khan, and S. Rehman, "A systematic literature review on municipal solid waste management using machine learning and deep learning," *Artificial Intelligence Review*, 2025.
- [17] Bigbelly, "Smart Trash Cans & Recycling Bins," 2026.
- [18] Sensoneo, "Smart Waste Management," 2026.
- [19] Ecube Labs, "Smart Waste Management Solutions," 2026.
- [20] Compology, "WasteOS smart waste monitoring and dynamic routing platform," 2026.

11.APPENDICES

Appendix A: Survey instruments

Stakeholder validation questionnaire that may be used if expert feedback is collected from municipal officers, waste collection staff, and recycling facility representatives during the refinement of the proposed system.

No.	Question	Response Format
1	How important is advance prediction of waste composition for recycling preparation in your organization?	5-point Likert scale
2	Which waste categories are most critical for operational planning?	Multiple choice
3	How often do unexpected waste composition changes affect your daily operations?	5-point Likert scale
4	Which contextual factors most influence waste composition in your area?	Multiple select
5	Would next-day forecasts improve staff or equipment allocation?	Yes / No + comments
6	What minimum prediction accuracy would make this system useful in practice?	Open-ended
7	Which dashboard views, charts, or alerts would be most valuable?	Open-ended
8	What concerns do you have regarding data quality, reliability, or adoption?	Open-ended

Appendix B: Risk analysis

Key project risks, their likely impact, and proposed mitigation actions.

Risk	Likelihood	Impact	Mitigation
Limited availability of real waste composition data	Medium	High	Use a combination of available records, public datasets, and carefully justified simulated data for prototype evaluation.
Poor data quality or missing values	High	Medium	Apply preprocessing, imputation, and data validation checks before model training.
Low model accuracy for certain waste categories	Medium	High	Compare multiple models, tune hyper parameters, and report MAE/RMSE/R ² transparently.
Weather API or external data access issues	Low	Medium	Cache retrieved data and prepare a fallback dataset for testing and demonstration.

Time constraints during development	Medium	Medium	Prioritize core prediction and dashboard features first, then extend optional functionality.
Stakeholder adoption challenges	Medium	Medium	Design a simple dashboard and include stakeholder validation feedback early in development.

Appendix C: Ethical documents

This study primarily uses environmental and operational data rather than personal or sensitive data. Therefore, the ethical risk level is low. The following ethical handling statements are proposed.

- **Data privacy** - No personally identifiable information will be intentionally collected, stored, or processed in the forecasting prototype.
- **Data security** - Project files, datasets, and model outputs will be stored in secured environments with controlled access.
- **Responsible data use** - Any public or third-party datasets will be used only for academic research purposes with proper acknowledgement.
- **Expert feedback** - If stakeholder interviews or questionnaires are conducted, participation will be voluntary and informed consent will be obtained.
- **Reporting transparency** - Model limitations, assumptions, and simulated data usage will be clearly disclosed in the final dissertation.