

E-Waste Prediction and Recycling Behavior Using Machine Learning

Vaishnavi Jondhale¹, Asmita Musale², Rehuti Sutar³

^{1,2,3}*Pillai College of Arts Commerce and Science Autonomous*

Abstract—Electronic waste (e-waste) is increasing due to the growing use, replacement, and disposal of electronic devices such as mobile phones, laptops, computers, televisions, and other electronic products. Improper disposal of these devices can result in environmental problems and loss of recyclable materials.

This research proposes a data-driven study to understand the factors associated with e-waste generation and recycling behaviour. Data will be collected from individuals or households using a structured questionnaire. The collected data will be analyzed using statistical techniques to identify important patterns and relationships.

Machine Learning techniques will be applied for two prediction tasks. The first task will focus on predicting the amount of e-waste generated based on factors such as electronic device ownership, replacement frequency, and disposal behaviour. The second task will focus on predicting recycling behaviour using factors such as awareness, recycling knowledge, accessibility, and previous recycling practices. Excel will be used for initial data organization and analysis, Python will be used for statistical analysis and Machine Learning, and MongoDB will be used for structured data storage. The study aims to demonstrate how predictive analytics can be applied to understand and predict e-waste generation and recycling behaviour.

Index Terms—E-Waste, Data Analytics, Recycling Behaviour, Machine Learning, Predictive Analytics, Statistical Analysis

I. INTRODUCTION

Electronic devices have become an essential part of modern life. Mobile phones, laptops, computers, televisions, tablets, chargers, and other electronic products are widely used by individuals and households.

As technology develops, people frequently replace older devices with newer models. This results in the generation of electronic waste, commonly known as e-

waste. E-waste contains different materials, including materials that can potentially be recovered and recycled. However, when electronic devices are discarded through improper methods, recyclable materials may be lost and environmental concerns may arise.

The generation of e-waste can depend on several factors, including the number of electronic devices owned, replacement frequency, device usage, and disposal practices. Similarly, recycling behaviour can be associated with awareness, knowledge, accessibility of recycling facilities, convenience, incentives, and previous recycling experience.

Therefore, studying these factors using data can help in understanding patterns related to e-waste generation and recycling behaviour.

This research proposes the use of statistical analysis and Machine Learning to study these patterns and develop prediction models.

The research will focus on two major prediction tasks:

1. Prediction of e-waste generation.
2. Prediction of recycling behaviour.

II. PROBLEM STATEMENT

The increasing use and replacement of electronic devices contributes to the generation of e-waste. At the same time, electronic waste is not always disposed of through appropriate recycling channels.

Different individuals and households may have different levels of electronic device usage, awareness, recycling knowledge, and access to recycling facilities. These factors may influence both the quantity of e-waste generated and recycling behaviour. Simple observation may not be sufficient to understand the relationship between these factors.

Therefore, this research aims to study the collected data using statistical analysis and Machine Learning.

Research Problem

How can data analytics and Machine Learning be used to identify factors associated with e-waste generation and recycling behaviour and predict future e-waste generation and recycling behaviour?

III. RESEARCH QUESTIONS

The research will attempt to answer the following questions:

1. What factors are associated with e-waste generation among individuals or households?
2. What factors are associated with recycling behaviour?
3. Is awareness about e-waste recycling associated with recycling behaviour?
4. Can e-waste generation be predicted using collected data?
5. Can recycling behaviour be predicted using Machine Learning?
6. Which factors contribute most to the prediction models?

IV. OBJECTIVES

The main objectives of this research are:

1. To collect data related to electronic device ownership, usage, disposal, and recycling behaviour.
2. To analyze factors associated with e-waste generation.
3. To study factors associated with recycling behaviour.
4. To apply statistical techniques to identify patterns and relationships within the collected data.
5. To develop a Machine Learning model for predicting e-waste generation.
6. To develop a Machine Learning model for predicting recycling behaviour.
7. To evaluate the performance of the developed prediction models.
8. To identify important factors contributing to the prediction of e-waste generation and recycling behaviour.

V. SIGNIFICANCE OF THE STUDY

The study applies data analytics and Machine Learning to an environmental problem.

The research can help in:

- Understanding patterns in e-waste generation.
- Identifying factors associated with recycling behaviour.
- Studying the role of awareness and recycling knowledge.
- Estimating e-waste generation using available data.
- Understanding factors that may be useful for predicting recycling behaviour.
- Demonstrating the application of statistical analysis and Machine Learning in e-waste research.

The study can also provide a foundation for future research using larger datasets and additional variables.

VI. LITERATURE REVIEW

Previous research on e-waste has examined areas such as e-waste generation, electronic device disposal, recycling practices, environmental awareness, and recycling behaviour.

Studies in this area have considered factors such as electronic device ownership, consumption patterns, device replacement, and useful life of electronic products when studying e-waste generation.

Research related to recycling behaviour has also examined factors such as environmental awareness, knowledge about recycling, availability of collection facilities, convenience, and previous recycling experience.

Statistical techniques can be used to summarize data and identify relationships between variables. Machine Learning techniques can additionally be used to develop predictive models from multiple input factors. Based on these areas of research, the proposed study will combine statistical analysis and Machine Learning to study both e-waste generation and recycling behaviour.

A detailed literature review containing relevant published research papers will be carried out before preparation of the final research paper.

VII. PROPOSED METHODOLOGY

The proposed research will follow the following methodology:

Data Collection → Data Organization → Data Cleaning → Statistical Analysis → Feature Selection → Machine Learning → Model Evaluation → Prediction → Findings

7.1. Data Collection

Primary data will be collected using a structured questionnaire. The questionnaire will collect information related to:

- Electronic devices owned
- Number of devices
- Device replacement frequency
- Devices discarded
- E-waste awareness
- Recycling knowledge
- Previous recycling experience
- Disposal methods
- Accessibility of recycling facilities
- Convenience of recycling
- Incentive awareness
- Recycling behaviour

The questionnaire may be distributed through an online form.

A proposed target of approximately 150 respondents will be considered for the initial study, subject to actual data collection and research requirements.

VIII. PROPOSED DATASET AND VARIABLES

The collected dataset may contain variables such as:

Variable	Description
Respondent_ID	Unique identification number
Age_Group	Age category
Household_Size	Number of household members
Device_Count	Number of electronic devices owned
Mobile_Count	Number of mobile phones
Laptop_Count	Number of laptops
Replacement_Frequency	Frequency of replacing devices
Devices_Discarded	Number of devices discarded
Estimated_EWaste_kg	Estimated amount of e-waste
Awareness_Score	Awareness about e-waste
Recycling_Knowledge	Knowledge about recycling
Recycling_Access	Access to recycling facilities
Recycling_Convenience	Convenience of recycling
Incentive_Awareness	Awareness of recycling incentives
Previous_Recycling	Previous recycling experience
Disposal_Method	Method used to dispose electronic devices
Willingness_to_Recycle	Willingness to recycle
Recycling_Behaviour	Recycling behaviour

The final variables will be finalized after designing and testing the questionnaire.

IX. DATA STORAGE AND PREPARATION

The collected responses will initially be organized in Excel.

The structured records can also be stored in MongoDB for data management. Before analysis, the dataset will be checked for:

- Missing values
- Duplicate records
- Incorrect entries
- Inconsistent formats
- Invalid values

Categorical variables will be converted into suitable formats when required for statistical analysis and Machine Learning.

X. STATISTICAL ANALYSIS

Statistical analysis will be performed before applying Machine Learning.

10.1.Descriptive Statistics

The following measures will be used where applicable:

- Mean
- Median
- Mode
- Minimum and maximum
- Standard deviation
- Frequency
- Percentage

These measures will help summarize the collected responses.

10.2.Correlation Analysis

Correlation analysis will be used to examine relationships between numerical variables.

For example:

Number of electronic devices ↔ Estimated e-waste and

Awareness score ↔ Recycling-related behaviour

10.3.Association Analysis

Where suitable, categorical variables will be analyzed to examine associations such as:

- Awareness level and recycling behaviour

- Disposal method and recycling behaviour
- Recycling access and recycling behaviour

The final statistical tests will be selected according to the collected data and variable types.

XI. MACHINE LEARNING APPROACH

Machine Learning will be applied to two major prediction tasks.

11.1.E-Waste Generation Prediction

The first task will predict the estimated quantity of e-waste generated.

Target Variable:

Estimated_EWaste_kg

Possible input variables include:

- Device count
- Devices discarded
- Replacement frequency
- Household size
- Device types
- Usage-related information

Since the target is numerical, this will be treated as a regression problem.

Possible algorithms:

- Linear Regression
- Decision Tree Regressor
- Random Forest Regressor

11.2.Recycling Behaviour Prediction

The second task will predict recycling behaviour. For example, the behaviour can be represented as:

Recycles / Does Not Recycle

based on the final questionnaire and dataset.

Possible input variables include:

- Awareness score
- Recycling knowledge
- Recycling access
- Recycling convenience
- Previous recycling experience
- Incentive awareness
- Disposal method
- Willingness to recycle

This will be treated as a classification problem.

Possible algorithms:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier

XII. MACHINE LEARNING WORKFLOW

The proposed Machine Learning process is:

Collected Data

↓

Data Cleaning

↓

Feature Selection

↓

Train-Test Split

↓

Model Training

↓

Prediction

↓

Model Evaluation

↓

Model Comparison

↓

Final Prediction

The collected dataset will be divided into training and testing portions.

The training data will be used to develop the models, while the testing data will be used to evaluate their performance.

XIII. MODEL EVALUATION

Different evaluation measures will be used for the two prediction tasks.

For E-Waste Generation Prediction

The following metrics may be used:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- R² Score

These metrics will help measure the difference between actual and predicted e-waste quantity.

For Recycling Behaviour Prediction

The following metrics may be used:

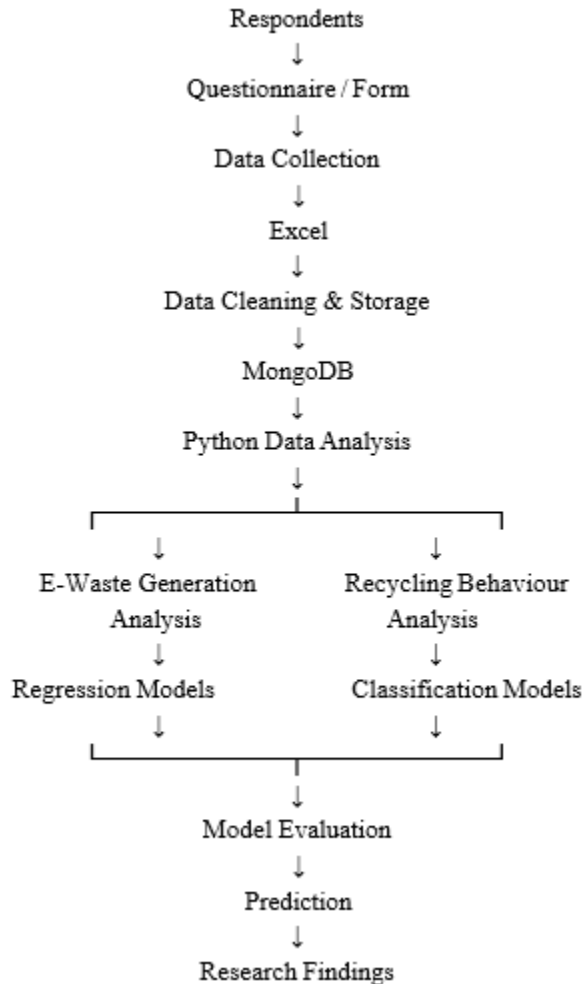
- Accuracy
- Precision

- Recall
- F1-Score
- Confusion Matrix

These metrics will help evaluate the classification performance of the models.

XIV. PROPOSED RESEARCH ARCHITECTURE

The overall research process can be represented as:



XV. EXPECTED OUTCOMES

The actual results will be obtained after data collection and analysis. The study is expected to provide information about:

1. Factors associated with e-waste generation.
2. Factors associated with recycling behaviour.
3. Relationship between electronic device ownership and e-waste generation.
4. Relationship between awareness and recycling

behaviour.

5. Performance of different regression models.

6. Performance of different classification models.

7. Important features contributing to the prediction models.

No numerical results will be presented in this proposal before the actual data collection and analysis.

XVI. LIMITATIONS

The proposed study may have the following limitations:

1. The sample size may be limited.
2. The study may depend on self-reported information.
3. Estimated e-waste quantity may not always represent physically measured e-waste.
4. Recycling behaviour may differ across different locations and populations.
5. Machine Learning performance will depend on the quality and size of the collected dataset.
6. The findings may not represent the entire population.

XVII. FUTURE SCOPE

The research can be extended in the future by:

1. Collecting a larger dataset from different geographical areas.
2. Including actual e-waste collection records.
3. Adding more categories of electronic devices.
4. Studying e-waste trends over a longer period.
5. Including additional Machine Learning algorithms.
6. Developing an interactive e-waste analytics dashboard.
7. Studying recycling behaviour across different demographic groups.
8. Developing a real-time prediction system using continuously collected data.

VIII. CONCLUSION

This proposed research focuses on the analysis and prediction of e-waste generation and recycling behaviour using data analytics and Machine Learning. Data will be collected from individuals or households through a structured questionnaire. The collected data will first be analyzed using statistical techniques to understand patterns and relationships.

Machine Learning models will then be developed for two prediction tasks: e-waste generation prediction

and recycling behaviour prediction. Suitable evaluation metrics will be used to measure model performance.

The research aims to demonstrate how statistical analysis and Machine Learning can be used to understand e-waste-related behaviour and develop data-driven predictions.

The actual findings and numerical results will be added after completion of data collection and analysis.

REFERENCES

- [1] United Nations Institute for Training and Research (UNITAR), *Global E-waste Monitor*. [Online]. Available: <https://ewastemonitor.info/>
- [2] World Health Organization (WHO), “Electronic Waste and Health.” [Online]. Available:

[https://www.who.int/news-room/fact-sheets/detail/electronic-waste-\(e-waste\)](https://www.who.int/news-room/fact-sheets/detail/electronic-waste-(e-waste))

- [3] Ministry of Environment, Forest and Climate Change, Government of India, “Information and Reports Related to E-waste Management.” [Online]. Available: <https://moef.gov.in/>
- [4] Scikit-learn, “Machine Learning in Python: Scikit-learn Documentation.” [Online]. Available: <https://scikit-learn.org/stable/>
- [5] MongoDB, “MongoDB Documentation.” [Online]. Available: <https://www.mongodb.com/docs/>
- [6] Microsoft Support, “Excel Data Analysis and Statistical Functions.” [Online]. Available: <https://support.microsoft.com/>

PLAGARISM REPORT

