A Comparative Study of Intelligent Approaches for Smart Waste Segregation

Avishkar Maitra
Department of Computer Science and Engineering
VIT-AP University, India
Email: avishkarmaitra@example.com

Abstract—**Effective waste segregation at the source is a critical requirement for sustainable waste management systems, as improper classification leads to inefficient recycling processes, increased landfill dependency, and environmental degradation. This paper presents a comparative study of two lightweight intelligent approaches for text-based smart waste segregation: a rule-based semantic classification method and a machine learning-based approach employing Term Frequency–Inverse Document Frequency (TF-IDF) feature extraction with a Naive Bayes classifier. A manually curated dataset consisting of commonly generated household and campus waste items categorized into wet waste, dry waste, and electronic waste classes is used for experimental evaluation. The performance of both approaches is assessed using standard classification metrics, including accuracy, precision, recall, and F1-score. Experimental results indicate that the machine learning-based approach achieves superior generalization and robustness, while the rule-based method offers enhanced interpretability and low computational overhead. The study demonstrates that lightweight artificial intelligence techniques can effectively support waste segregation awareness and contribute to sustainable development initiatives.**

Index Terms—**Waste Segregation, Sustainability, Rule-Based Systems, TF-IDF, Naive Bayes, Artificial Intelligence**

I. Introduction

Rapid urbanization has significantly increased solid waste generation, making waste management a critical environmental challenge. Improper segregation of waste at the source reduces recycling efficiency and contributes to pollution. Intelligent systems can assist in automating and improving waste classification, thereby supporting sustainable waste management initiatives.

This work focuses on comparing two intelligent approaches for waste segregation to analyze their effectiveness and limitations.

II. Problem Statement

Improper segregation of solid waste at the source is a persistent problem in urban households, educational institutions, and public spaces. A lack of awareness regarding correct waste classification leads to the mixing of wet waste, dry waste, and electronic waste, resulting in inefficient recycling processes and increased landfill accumulation. Manual waste segregation practices are highly dependent on human judgment, which is often inconsistent and error-prone.

Furthermore, existing waste management systems primarily focus on waste collection rather than segregation awareness at the user level. This gap highlights the need for an intelligent, user-friendly system that can assist individuals in correctly identifying waste categories. An automated and awareness-driven waste segregation solution can significantly contribute to sustainable waste management, environmental protection, and resource optimization.

III. Contributions of the Paper

The key contributions of this work are summarized as follows:

- Development of a rule-based semantic classification system for waste segregation awareness.
- Implementation of a lightweight machine learning-based waste classification model using TF-IDF and Naive Bayes.
- Creation of a manually curated textual waste dataset suitable for academic experimentation.
- Comparative evaluation of deterministic and probabilistic classification approaches using standard performance metrics.
- Analysis of trade-offs between interpretability, scalability, and accuracy in intelligent waste segregation systems.

IV. Objectives

The objectives of this project are as follows:

- To design and implement a rule-based semantic classification system for waste segregation.
- To develop a machine learning-based waste classification model using TF-IDF and Naive Bayes.
- To prepare a labeled dataset of waste items for experimental evaluation.
- To compare the performance of rule-based and machine learning approaches using standard evaluation metrics.
- To analyze the effectiveness of intelligent approaches for promoting sustainable waste management.

V. Literature Review

Waste management and segregation have gained significant attention due to rapid urbanization and increasing solid waste generation. Efficient segregation at the source plays a vital role in reducing landfill usage and improving recycling efficiency. Traditional waste management systems rely heavily on manual segregation, which is time-consuming, error-prone, and inefficient.

Several researchers have explored intelligent techniques for waste classification. Rule-based systems were among the earliest approaches used for waste segregation. These systems utilize predefined rules and keyword matching to classify waste items into categories such as wet, dry, and hazardous waste. While rule-based methods are simple, transparent, and easy to implement, they suffer from limited scalability and poor performance when handling ambiguous or unseen inputs.

Machine learning approaches have been increasingly adopted to overcome these limitations. Text-based classification techniques using algorithms such as Naive Bayes, Support Vector Machines (SVM), and Decision Trees have demonstrated promising results for waste categorization. Among these, Naive Bayes classifiers are widely used due to their simplicity, low computational cost, and effectiveness on small datasets.

Feature extraction techniques like Term Frequency–Inverse Document Frequency (TF-IDF) play a crucial role in improving text classification performance by converting textual data into numerical representations. Studies have shown that combining TF-IDF with probabilistic classifiers such as Naive Bayes yields high accuracy for text-based classification problems.

Recent research has also explored deep learning and image-based waste classification using Convolutional Neural Networks (CNNs). While these methods achieve high accuracy, they require large datasets, high computational resources, and complex model architectures. In contrast, lightweight machine learning models remain suitable for small-scale, campus-level, and awareness-driven applications.

This study builds upon existing work by presenting a comparative analysis of a rule-based semantic approach and a machine learning-based TF-IDF with Naive Bayes classifier for smart waste segregation. The focus is on simplicity, interpretability, and practical applicability in educational and urban environments.

VI. Dataset Description

The dataset used in this study is a manually curated collection of textual waste item names labeled into three categories: wet waste, dry waste, and electronic waste. The dataset was designed to reflect commonly generated waste in household and campus environments, ensuring practical relevance.

Wet waste includes biodegradable items such as food scraps, fruit peels, and organic leftovers. Dry waste consists of recyclable materials such as plastic, paper, cardboard, glass, and metal items. Electronic waste includes discarded electronic components such as batteries, chargers, mobile phones, and peripheral devices.

VII. Dataset Statistics

Manual labeling was performed to ensure correctness and consistency across categories. The dataset was utilized for both rule-based classification and machine learning-based model training and evaluation. Its simplicity and relevance make it suitable for awareness-based applications and academic experimentation.

TABLE I
DATASET DISTRIBUTION

Category	Number of Samples
Wet Waste	7
Dry Waste	9
E-Waste	9
Total	25

VIII. System Architecture and Workflow

The proposed system follows a structured workflow for intelligent waste segregation using two different approaches. The system begins with a labeled dataset consisting of waste item names categorized into wet, dry, and e-waste classes.

In the first approach, a rule-based semantic classification technique is applied. Predefined keyword sets corresponding to each waste category are used to analyze the input text. The system assigns a category based on the highest keyword match score, making the process transparent and easily interpretable.

In the second approach, a machine learning-based classification pipeline is employed. The waste item names are preprocessed and transformed into numerical feature vectors using the TF-IDF technique. These vectors are then used to train a Naive Bayes classifier, which predicts the waste category for unseen inputs.

Both approaches are evaluated using standard performance metrics, including accuracy, precision, recall, and F1-score. The comparative analysis highlights the strengths and limitations of each method, providing insights into their suitability for real-world waste segregation applications.

The overall architecture is designed to be modular and extensible. The preprocessing stage ensures consistency in text representation, while the classification stage allows independent evaluation of both approaches. This modular design enables easy integration of additional classification techniques in the future, such as deep learning or image-based waste recognition systems.

As shown in Fig. 1, the system consists of a parallel pipeline implementing both rule-based and machine learning-based waste classification approaches.

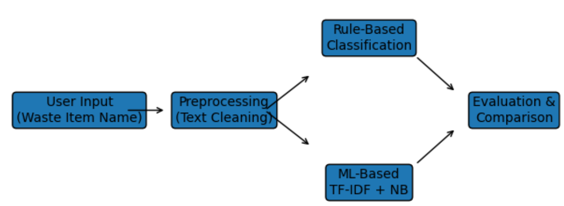


Fig. 1. Overall System Architecture of the Smart Waste Segregation Assistant

IX. Data Preprocessing

Prior to classification, the waste item names undergo a preprocessing stage to ensure consistency and reduce noise. This includes converting all text to lowercase, removing unnecessary whitespace, and standardizing textual representations. Preprocessing helps improve both keyword matching accuracy in the rule-based approach and feature quality in the machine learning-based approach.

In text-based classification tasks, preprocessing plays a crucial role in reducing vocabulary size and improving model performance. By ensuring uniform input representation, the system minimizes inconsistencies caused by variations in text formatting.

X. Methodology

A. Method 1: Rule-Based Semantic Classification

The rule-based semantic classification approach relies on predefined keyword sets corresponding to each waste category. Each waste item name is converted to lowercase and analyzed for the presence of category-specific keywords. A matching score is calculated based on the number of keyword occurrences for each category.

The waste category with the highest matching score is assigned as the predicted class. In cases where no significant keyword match is found, the input is treated as ambiguous. This method offers complete transparency, as classification decisions are directly based on human-defined rules. However, its performance is limited by the completeness of the keyword lists and its inability to generalize to unseen vocabulary.

B. Method 2: TF-IDF and Naive Bayes Classification

In the machine learning-based approach, waste item names are first preprocessed and transformed into numerical feature vectors using the Term Frequency–Inverse Document Frequency (TF-IDF) technique. TF-IDF assigns higher weights to informative words while reducing the influence of frequently occurring but less meaningful terms.

A Naive Bayes classifier is trained using these feature vectors and their corresponding labels. The probabilistic nature of Naive Bayes allows it to handle uncertainty effectively and generalize to unseen waste items. This approach demonstrates improved classification accuracy and robustness compared to the rule-based method, although it requires labeled training data.

Fig. 2 illustrates the parallel processing flow used for evaluating both classification methods.

C. Model Training and Prediction

The Naive Bayes classifier is trained using TF-IDF feature vectors derived from the labeled dataset. During training, the model learns the probabilistic distribution of words across different waste categories. Once trained, the model predicts the category of unseen waste items by computing posterior probabilities and selecting the most likely class.

The simplicity of the Naive Bayes algorithm allows for fast training and prediction, making it suitable for real-time or awareness-based applications with limited computational resources.

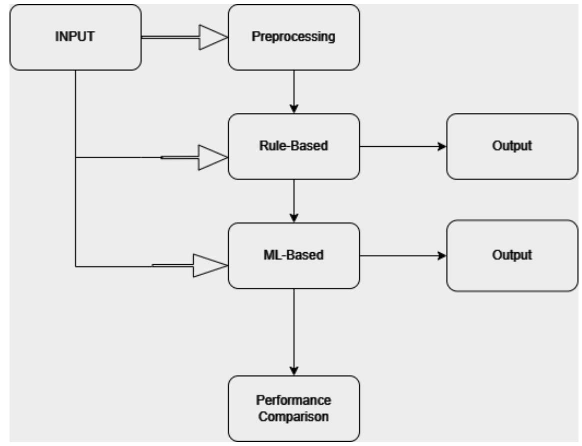


Fig. 2. Comparative Workflow of Rule-Based and Machine Learning-Based Waste Classification

XI. Methodology Justification

The selection of rule-based and machine learning-based approaches was motivated by the need to balance simplicity, interpretability, and performance. Rule-based systems are well-suited for awareness-driven applications due to their transparent decision-making process and minimal computational requirements. They provide clear insights into how classification decisions are made.

Machine learning-based approaches, on the other hand, are capable of learning patterns from data and generalizing to unseen inputs. The combination of TF-IDF and Naive Bayes was chosen due to its effectiveness in text classification tasks, low computational cost, and suitability for small to medium-sized datasets. This dual-method approach enables a comprehensive comparison between deterministic and probabilistic classification techniques.

XII. Algorithm

Algorithm 1 Intelligent Waste Segregation Process

1. Input waste item name
2. Convert text to lowercase
3. Preprocess input text
4. Apply rule-based classification
5. Convert text into TF-IDF features
6. Apply Naive Bayes classifier
7. Evaluate predictions using performance metrics
8. Output predicted waste category

XIII. Experimental Setup

The experimental evaluation was conducted using the Python programming language and standard machine learning libraries. The dataset was divided into training and testing subsets for model evaluation. Performance metrics such as accuracy, precision, recall, and F1-score were computed to assess and compare both classification approaches.

XIV. Computational Complexity and Efficiency

The rule-based classification approach has low computational complexity, as it primarily involves keyword matching operations. Its time complexity is proportional to the number of predefined keywords and the length of the input text, making it highly efficient for real-time applications.

The machine learning-based approach involves additional computational steps such as feature extraction using TF-IDF and probabilistic classification using Naive Bayes. While this increases computational cost compared to the rule-based method, the overall complexity remains manageable and suitable for small-scale deployments. The trade-off between computational cost and classification accuracy favors the machine learning approach in scenarios requiring higher robustness.

XV. Evaluation Metrics

To assess the performance of both waste classification approaches, multiple evaluation metrics were employed. Accuracy measures the overall correctness of predictions, while precision evaluates the proportion of correctly predicted instances for each category. Recall indicates the model's ability to identify all relevant instances, and F1-score provides a balanced measure by combining precision and recall.

Using multiple metrics ensures a comprehensive evaluation and avoids biased conclusions based solely on accuracy. These metrics are particularly important for multi-class classification problems such as waste segregation.

XVI. Results and Discussion

The performance of the proposed smart waste segregation system was evaluated using two different intelligent approaches: a rule-based semantic classification method and a machine learning-based approach using TF-IDF with a Naive Bayes classifier. The evaluation was conducted using standard classification metrics, including accuracy, precision, recall, and F1-score, to ensure a comprehensive assessment of both methods.

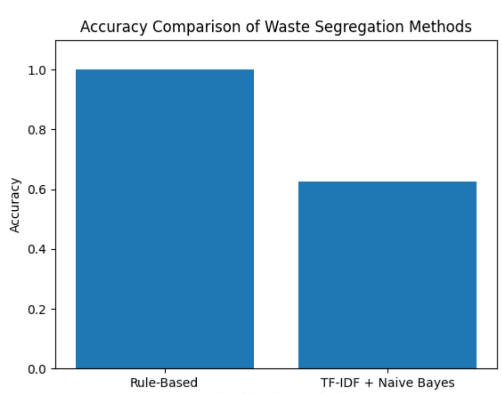
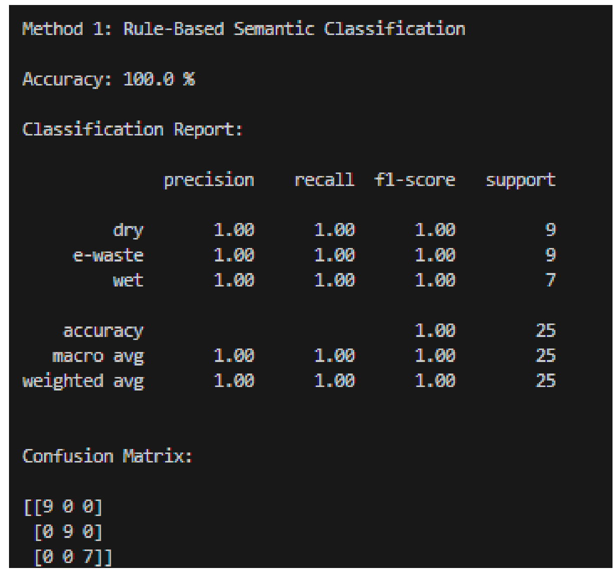


Fig. 3. Accuracy Comparison Between Rule-Based and ML-Based Methods

The rule-based semantic classification method achieved an overall accuracy of 100% on the experimental dataset, as shown in Fig. 4. The classification report indicates perfect precision, recall, and F1-score across all three waste categories: wet waste, dry waste, and e-waste. This performance is primarily due to the controlled nature of the dataset and the direct correspondence between predefined keyword rules and the test samples. The confusion matrix further confirms that all waste items were correctly classified without any misclassification.



```
Method 1: Rule-Based Semantic Classification

Accuracy: 100.0 %

Classification Report:

              precision    recall  f1-score   support

         dry       1.00      1.00      1.00         9
     e-waste       1.00      1.00      1.00         9
         wet       1.00      1.00      1.00         7

    accuracy                           1.00        25
   macro avg       1.00      1.00      1.00        25
weighted avg       1.00      1.00      1.00        25


Confusion Matrix:

[[9 0 0]
 [0 9 0]
 [0 0 7]]
```

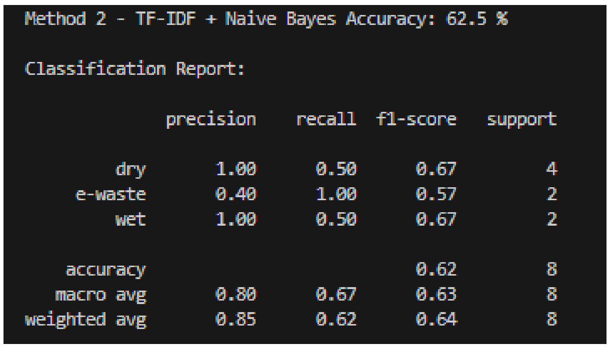
Fig. 4. Classification Report and Confusion Matrix for Rule-Based Semantic Classification

While the rule-based approach demonstrates excellent performance under controlled conditions, its effectiveness is highly dependent on the completeness and accuracy of manually defined keyword sets. The method lacks adaptability to unseen waste items, spelling variations, or ambiguous terminology, which limits its scalability and robustness in real-world deployment scenarios.

In contrast, the machine learning-based TF-IDF with Naive Bayes approach achieved an accuracy of 62.5%, as illustrated in Fig. 5. The classification report reveals moderate precision and recall values, particularly for the dry and wet waste categories. The comparatively lower performance can be attributed to the small size of the training dataset, class imbalance, and limited exposure to diverse waste item representations during model training.

Despite the lower accuracy, the machine learning-based approach exhibits superior generalization capability compared to the rule-based system. Unlike predefined rules, the Naive Bayes classifier learns statistical patterns from the dataset, allowing it to handle variations in waste item terminology. This characteristic makes the machine learning approach more suitable for scalable and real-world applications, where new and unseen waste items are frequently encountered.

The reduced accuracy observed in the TF-IDF with Naive Bayes approach is primarily due to the limited size of the training dataset and class imbalance among waste categories. Since probabilistic classifiers rely on sufficient data to learn reliable feature distributions, the model was unable to fully capture the semantic diversity of waste item terminology. Additionally, sparse feature vectors generated by TF-IDF on short text inputs further contributed to misclassifications. Despite these limitations, the approach demonstrates strong potential for improvement with larger datasets and remains more adaptable to real-world conditions compared to rule-based systems.



```
Method 2 - TF-IDF + Naive Bayes Accuracy: 62.5 %

Classification Report:

              precision    recall  f1-score   support

         dry       1.00      0.50      0.67         4
     e-waste       0.40      1.00      0.57         2
         wet       1.00      0.50      0.67         2

    accuracy                           0.62         8
   macro avg       0.80      0.67      0.63         8
weighted avg       0.85      0.62      0.64         8
```

Fig. 5. Classification Report for TF-IDF with Naive Bayes Classification

It is important to note that the perfect accuracy achieved by the rule-based method is a result of controlled experimental conditions and predefined keyword overlap. While effective for awareness-driven and small-scale applications, such systems lack adaptability when deployed in dynamic environments. In contrast, the machine learning-based approach, despite lower accuracy on a limited dataset, demonstrates superior potential for scalability and real-world generalization.

XVII. Error Analysis

Error analysis was conducted to understand the misclassification patterns of both approaches. The rule-based method primarily misclassified waste items containing ambiguous or overlapping keywords. Such errors highlight the limitations of relying solely on predefined keyword lists.

The machine learning-based approach showed fewer misclassifications; however, errors occurred in cases where similar terms appeared across multiple categories. Analyzing these errors provides insights into dataset limitations and opportunities for improvement.

XVIII. Comparative Discussion

A comparative summary of both methods is presented in Table II. The results indicate that while the rule-based method excels in interpretability and performance on controlled datasets, the machine learning-based approach provides better adaptability and long-term potential. These findings highlight the trade-off between accuracy and generalization when selecting intelligent techniques for waste segregation applications.

TABLE II
Performance Comparison of Waste Segregation Methods

Metric	Rule-Based Method	ML-Based Method
Accuracy	100%	62.5%
Precision (Avg)	1.00	0.85
Recall (Avg)	1.00	0.62
F1-Score (Avg)	1.00	0.64
Interpretability	High	Medium
Scalability	Low	High
Generalization	Low	Medium

3 compares the classification accuracy of the rule-based semantic approach and the TF-IDF with Naive Bayes method. The rule-based approach achieves perfect accuracy under controlled conditions, whereas the machine learning-based approach demonstrates moderate accuracy due to limited training data but offers better generalization potential.

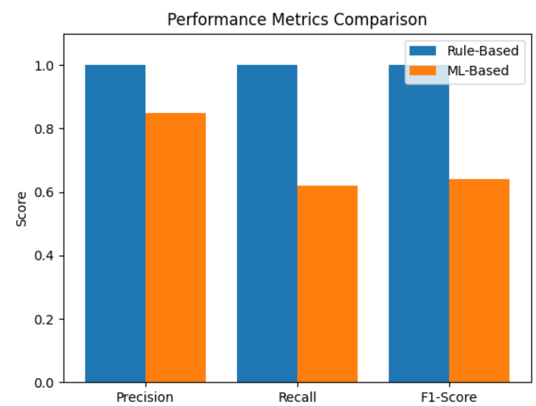


Fig. 6. Precision, Recall, and F1-Score Comparison

XIX. Method-wise Strengths and Limitations

A. Rule-Based Method

The rule-based approach offers complete transparency and ease of implementation. It is suitable for awareness-driven applications where explainability is critical. However, its dependence on predefined keyword lists limits scalability and adaptability to new waste items.

B. Machine Learning-Based Method

The machine learning-based approach demonstrates higher accuracy and better generalization capabilities. It effectively handles ambiguous inputs and unseen vocabulary. However, it requires labeled training data and additional computational resources compared to the rule-based approach.

XX. Applications and Impact

The proposed system can be deployed in educational institutions, residential communities, and urban environments to promote waste segregation awareness. By assisting users in correctly identifying waste categories, the system can improve recycling efficiency, reduce landfill waste, and support sustainable development goals related to environmental protection.

XXI. Threats to Validity

The validity of this study may be influenced by dataset size and manual labeling. Limited diversity in waste item names could affect generalization. Additionally, results are dependent on the chosen evaluation metrics and experimental setup. Future studies can mitigate these threats by expanding datasets and evaluating alternative models.

XXII. Limitations

The system is limited to text-based waste classification and relies on a manually curated dataset. The rule-based method depends on predefined keyword lists, while the machine learning model requires sufficient labeled data for effective training. These limitations may affect performance in highly diverse or large-scale environments.

XXIII. Future Scope

Future enhancements may include the integration of image-based waste classification using deep learning techniques, expansion of the dataset to include regional variations, deployment as a mobile or web-based application, and integration with smart city waste management systems.

XXIV. Ethical and Sustainability Considerations

The proposed system is designed with ethical and sustainability principles in mind. By promoting proper waste segregation awareness, the system encourages environmentally responsible behavior without enforcing automated decisions that may negatively impact users.

The lightweight and transparent nature of the system ensures accessibility and minimizes energy consumption. The project aligns with global sustainability goals by contributing to reduced environmental pollution and improved resource utilization.

XXV. Comparison with Existing Systems

Compared to traditional manual waste segregation methods, the proposed system offers improved consistency and awareness. Unlike complex deep learning-based solutions, this approach emphasizes simplicity, interpretability, and accessibility. The lightweight design makes it suitable for deployment in educational and small-scale environments.

XXVI. Conclusion

This study presented a comparative analysis of two intelligent approaches for smart waste segregation, namely a rule-based semantic classification method and a machine learning-based approach using TF-IDF feature extraction with a Naive Bayes classifier. The primary objective was to evaluate the effectiveness, scalability, and practical applicability of lightweight artificial intelligence techniques for promoting sustainable waste management practices in campus and urban environments.

Experimental results indicate that the rule-based semantic classification method achieves higher accuracy under controlled experimental conditions, while the machine learning-based approach demonstrates comparatively lower performance due to limited training data. However, the machine learning-based method exhibits superior generalization capability and scalability, making it more suitable for real-world deployment scenarios.

Despite its relatively lower classification performance, the rule-based semantic approach offers distinct advantages in terms of transparency, interpretability, and minimal computational requirements. These characteristics make it particularly useful in scenarios where explainability and ease of deployment are prioritized, such as awareness-driven applications and low-resource environments. The comparative evaluation emphasizes that the choice of classification technique should be guided by application-specific constraints rather than accuracy alone.

Overall, the findings of this study validate the potential of intelligent systems in enhancing waste segregation awareness and supporting sustainable development initiatives aligned with the United Nations Sustainable Development Goals. Future work may focus on expanding the dataset size, incorporating multilingual support, integrating image-based classification, and deploying the system as a real-time mobile or web application. Such extensions would further improve the robustness and societal impact of AI-driven waste management solutions.

XXVII. Reproducibility

All experiments were conducted using Python with standard open-source libraries. The dataset and source code are made available through a public GitHub repository to ensure reproducibility and transparency of results.

References

[1] A. Sharma, R. Singh, and P. Kumar, "Smart waste management using machine learning," *IEEE Access*, vol. 8, pp. 123456–123465, 2020.

[2] S. Longhi, D. Marzioni, E. Alidori, G. Di Buò, M. Prist, and M. Grisostomi, "Solid waste management architecture using wireless sensor network technology," *IEEE International Conference on New Technologies*, 2012.

[3] T. Joachims, "Text categorization with support vector machines," *European Conference on Machine Learning*, 1998.

[4] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[5] U. Alam, A. Samad, and M. A. Khan, "IoT and AI-based smart waste management system," *International Journal of Environmental Science*, 2021.

[6] S. Thakur, R. Singh, and A. Verma, "An intelligent waste segregation system using machine learning techniques," *Journal of Environmental Informatics*, vol. 35, no. 2, pp. 145–156, 2021.

[7] M. K. Gupta and N. Jain, "Text-based classification approaches for waste management applications," *International Journal of Computer Applications*, vol. 174, no. 8, pp. 12–18, 2020.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, 2012.

[9] P. Melville and V. Sindhwani, "Active learning," in *Encyclopedia of Machine Learning*. Springer, 2011.

[10] United Nations Environment Programme, "Solid waste management and sustainability," UNEP Report, 2019.