

Machine learning-based prediction model for high-value waste types to support the circular economy

Andi Irmayana¹, Nurul Aini², Ramlah P.³, Muhammad Husnul Yaqin¹, Zalsabila Sabir¹

¹Department of Information Systems, Universitas DIPA Makassar, Makassar, Indonesia

²Department of Software Engineering, Universitas DIPA Makassar, Makassar, Indonesia

³Department of Digital Business, Universitas Teknologi Akba Makassar, Makassar, Indonesia

Article Info

Article history:

Received Sep 30, 2025

Revised Jul 16, 2026

Accepted Aug 12, 2026

Keywords:

Circular economy

High value

Machine learning

Prediction

Waste

ABSTRACT

The management of high-value waste is a strategic challenge in supporting the implementation of a circular economy at the city level. This study aims to develop and compare machine learning-based prediction models to classify the economic value of waste into low, medium, and high categories based on the characteristics of transaction types and quantities. The dataset was obtained from waste bank transactions in Makassar City from January 2023 to June 2025. The research stages included data preprocessing, category labeling using the k-means and quantile-based classification algorithms, and then creating classification modeling using random forest and gradient boosting. The evaluation results show that both models perform very well with an accuracy rate above 92%, where gradient boosting consistently shows superior performance in terms of F1-score and the ability to reduce classification errors in high-value categories. Confusion matrix analysis shows that the boosting approach is more adaptive in handling class boundary distributions, while random forest tends to be more stable against data variations. These findings confirm that the ensemble learning approach is effective in supporting the classification of waste economic value based on historical transaction data.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Andi Irmayana

Department of Information Systems, Universitas DIPA Makassar

Perintis Kemerdekaan St. Km. 9, Tamalanrea, Makassar, Indonesia

Email: irmayana.andi@undipa.ac.id

1. INTRODUCTION

Waste management is a major challenge faced by many cities in Indonesia, including Makassar. Based on data from the Makassar City Environment Agency [1], waste generation in 2023 will reach more than 900 tons per day, with the majority coming from the household sector. Most of this waste is still not managed optimally, causing various environmental problems such as pollution, limited landfill space, and low recycling effectiveness. This phenomenon highlights the need for innovation in a more sustainable waste management system based on the principles of a circular economy. Plastic and mixed plastic waste are now in the global spotlight due to their increasing production rates, their difficulty in decomposing over long periods of time, and the continued dominance of environmentally unfriendly management practices. Therefore, sustainable plastic waste management through recycling and upcycling approaches is crucial to support the achievement of the sustainable development goal (SDG) [2]. When waste is mixed with feces or other organic materials, it is quality and value decrease drastically [3]. This condition results in the loss of many economic opportunities from household waste and hinders community participation in supporting the circular economy.

Factors such as public knowledge about waste management, facility support, household awareness levels, and the role of the government have a real and significant impact on public participation in waste management [1]. One of the fundamental problems is the low level of public knowledge, especially among housewives, about the economic value of waste [4]. Various types of waste, such as plastic, paper, metal, glass bottles, and used cooking oil, are often disposed of directly with organic waste without sorting. The authors [5]–[8] shows that plastic waste has a fairly high selling value and can be processed into new industrial raw materials. This context is also relevant to the conditions in Makassar City, where plastic waste dominates waste generation and has the potential to provide economic value if managed properly [9]. Not only plastic waste, but various other types of waste also have high economic value if managed properly through recycling processes. The authors [10]–[12] shows that the utilization of non-plastic waste through recycling technology is not only able to reduce the volume of waste accumulation, but also contributes significantly to increasing community income and environmental sustainability. This proves that integrated waste management, which is not limited to plastic, can be an effective strategy in supporting the circular economy and encouraging the growth of environmentally-based micro, small, and medium enterprises (MSMEs).

The focus of this research is to design a machine learning-based prediction system capable of identifying high-value waste types so that they can be sorted at the source. This approach is directed at historical waste collection data from waste banks in Makassar City, which includes type, volume, collection time, and selling price. By utilizing machine learning technology, waste value trends can be studied systematically to produce accurate recommendations. The two main algorithms that will be used in this research are random forest and gradient boosting, which are known to perform well in classification and data value prediction [13], [14].

The random forest algorithm was developed by Leo Breiman [15]. Random forest is one of the most powerful ensemble-based classification methods and is widely used in various machine learning applications [16]. Random forest works by forming a number of decision trees from randomly selected data subsets and also randomly selecting features at each split node, then combining the prediction results of each tree through majority voting for classification or averaging for regression [17]. This technique is effective in handling data with complex characteristics and provides a feature importance measure that shows how much each feature contributes to the prediction [18], [19]. Meanwhile, gradient boosting was introduced by Jerome H. Friedman and has since been further developed and applied in numerous recent studies [20], which compiles prediction models sequentially (boosting), where each subsequent model attempts to correct the errors of the previous model [21], resulting in highly accurate predictions, especially for non-linear and high-noise data [22]. The synergy between these two approaches opens up opportunities to build prediction systems that are not only precise, but also stable and flexible in the face of data variations in the field. Therefore, this study focuses on the design and comparison of random forest and gradient boosting prediction models to identify waste economic value categories more systematically. The main contribution of this study lies in the evaluation of the performance of both ensemble learning approaches in the context of waste bank transaction data, thereby providing an empirical basis for the development of data-driven decision-making systems in urban waste management.

2. RELATED WORKS

The circular economy is a sustainable development paradigm that emphasizes the optimization of resource utilization through the principles of reduce, reuse, and recycle. In the context of waste management, this concept aims to reduce dependence on primary resources by maximizing the value of used materials. With the development of technology, machine learning has been widely applied in waste management due to its ability to recognize complex patterns, perform classification, and predict material characteristics. Various algorithms, such as support vector machine (SVM), random forest, and deep learning approaches, have been proven to improve waste classification accuracy while reducing operational costs and accelerating the recycling process. Ranjbar *et al.* [23] shows that deep learning technology can effectively identify plastic types from construction waste based on RGB images. This finding emphasizes the importance of accurate material classification as the foundation for recycling optimization. Yevle and Mann [24] also highlights the development of AI in waste management, including data challenges, system integration, and implementation issues in the field. Meanwhile, Nahiduzzaman *et al.* [25] shows that deep learning-based waste classification can improve the efficiency of the recycling process and support environmental sustainability. A comprehensive study in [26] maps the application of machine learning and deep learning in urban solid waste management and confirms that although the technological foundation is strong, there is still room for development, particularly in terms of predicting the economic value of materials.

The application of machine learning for urban waste management is growing. Hidayah *et al.* [27] uses a clustering approach to group environmental indicators between provinces, demonstrating the relevance of unsupervised learning methods in the context of national environmental data. Meanwhile, Magriaty *et al.* [28] applies k-means to identify waste management areas in South Kalimantan and proves the effectiveness of clustering in analyzing regional waste characteristics. Another study, compares random forest and gradient boosting in environmental data classification and shows that the ensemble approach provides stable performance and good generalization capabilities on heterogeneous datasets [29]. The k-means method is widely used for initial waste data clustering, particularly in the operational sorting stage and waste production characteristic analysis [27], [28]. This approach enables the identification of natural patterns in data distribution without requiring initial labels, making it relevant for use as a pseudo-labeling mechanism in transaction datasets that do not have economic value categories.

To build a more robust prediction model in the context of environmental and transaction data, various studies have utilized ensemble learning approaches, particularly random forest and gradient boosting. Random forest works by building a number of decision trees in parallel and combining the prediction results through a majority voting mechanism. This approach is capable of handling non-linear relationships, reducing model variance, and maintaining prediction stability in data containing noise or heterogeneity [30]. The implementation of random forest in the classification of waste types based on digital images in Indonesia shows that this method can achieve high accuracy and provide strong interpretability through feature importance analysis [31]. In addition, a study on water quality prediction in Indonesia compared the performance of random forest and gradient boosting, where random forest showed competitive performance in real environmental classification evaluations [32].

Meanwhile, gradient boosting builds models sequentially (sequential learning) where each new model is designed to correct the errors of the previous model. This approach produces more adaptive decision boundaries and is able to systematically minimize residual error, thus demonstrating high generalization capabilities on complex data [32]. These two methods were chosen to improve generalization, handle noise, and provide model interpretability, while supporting the prediction of high-value waste in order to strengthen the concept of a circular economy. Although various studies have applied machine learning to waste type classification and waste management optimization, studies that specifically focus on predicting high economic value waste based on real transaction data are still limited. The novelty of this research lies in the integration of clustering-based labeling and quantile classification strategies to produce objective economic value categories, as well as the application of an ensemble learning approach through random forest and gradient boosting, reinforced with stratified cross-validation to improve model robustness. Unlike previous studies that focused on general classification, this research explicitly links predictive modeling with economic valorization strategies within the circular economy framework. By utilizing waste bank transaction data, this research provides methodological and practical contributions in supporting data-driven decision-making for material recovery optimization and strengthening the circular economy system at the local level.

3. METHOD

3.1. Research design

This study uses an experimental quantitative approach with a supervised classification-based machine learning method. This approach was chosen to support decision-making in the management of economically valuable waste and to support the implementation of a data-based circular economy. Secondary data was obtained from the records of the Makassar City waste bank, which sourced waste purchases from several waste bank units. The main focus was to build a prediction model for high-value waste types using the random forest and gradient boosting algorithms. Random forest is a machine learning technique that combines several decision trees to reduce correlations between data features and mitigate correlations between decision trees by using random sampling and feature selection [33]. Gradient boosting is a supervised machine learning ensemble technique that sequentially combines several weak models to form a stronger, more accurate model [22]. The research design includes data collection, dataset exploration, preprocessing, clustering to obtain labels, modelling with machine learning algorithms, and model performance evaluation. This study uses a supervised learning framework, where historical data (type of waste, volume, and price) is used as training data to predict high-value waste categories.

3.2. Research procedure

Each stage is designed to ensure the resulting model has a robust methodological foundation. At the same time, the model remains practically applicable to support decision-making regarding waste banks. The overall step-by-step workflow is illustrated in detail in Figure 1.

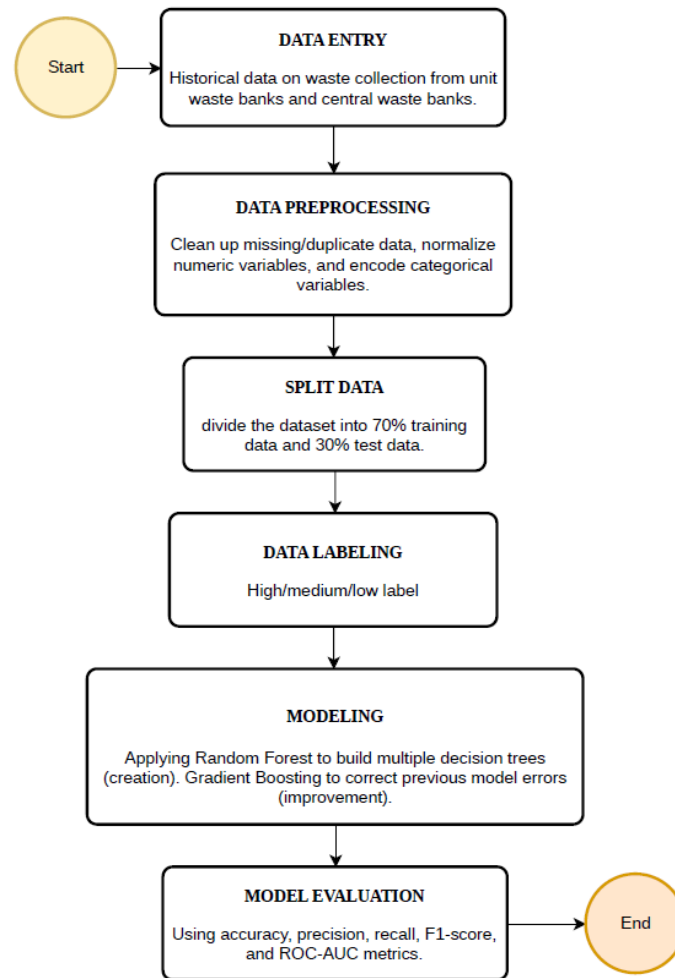


Figure 1. Research algorithm

Based on Figure 1, the research procedure consists of several stages, namely:

- i) The input data consists of historical data on waste purchases from unit waste banks to the central waste bank in Makassar, covering 45,544 data points after the cleaning process. The main attributes in the dataset include: i) codes: waste codes or types; ii) quantity: amount of waste traded; and iii) value: economic value of the transaction (Rp). Transaction value data is used only for labeling purposes and is not used as model input features to avoid data leakage.
- ii) Data pre-processing includes: i) data cleaning, by removing data with empty or invalid values; ii) data type conversion, specifically converting the quantity attribute from a non-numeric type to a numeric type; iii) numeric data normalization, using StandardScaler to maintain model training stability; and iv) categorical data encoding, using label encoding on the codes attribute.
- iii) Data labeling, due to the unavailability of labels in the dataset, a labeling mechanism is needed to produce objective and methodologically accountable economic value categories. Therefore, at this stage, two labeling strategies are applied as follows:
 - Clustering-based labeling (k-means): initial labeling was performed by grouping data based on volume attributes (quantity) and transaction values (value) using the k-means algorithm. The clustering results were then mapped into three categories: high, medium, and low. This approach aims to capture natural patterns in data distribution. Nugraha *et al.* [34] shows that the k-means algorithm is effective in improving waste management efficiency through grouping based on specific characteristics.
 - Quantile-based classification: to ensure objectivity and reduce the influence of outliers, economic value labeling was also carried out using a quantile-based approach to the distribution of transaction values [35]. This method allows for proportional categorization based on the statistical characteristics of the data. The criteria for labeling economic values are determined as follows:

- low economic value: value $\leq$ 25% quantile (Q1), medium economic value: $Q1 < \text{value} \leq$ 75th quantile (Q3), and high economic value: value $>$ 75th quantile (Q3).
- iv) Split the data, dividing the dataset into 70% training data and 30% test data. The division is performed using stratified train-test split to keep the class proportions in the training and test data consistent [36].
 - v) Models and algorithms: this study uses two main classification algorithms, namely random forest, which is used as the main model due to its robustness, stability, and ability to reduce the risk of overfitting. gradient boosting is used as a comparative model to evaluate alternative performances.
 - vi) Model evaluation is performed using a confusion matrix, classification report, and F1-score (macro average). In addition, stratified k-fold cross validation (5-fold) is used to measure the model's generalization ability and ensure that the model does not experience overfitting [37], [38].

3.3. Data acquisition

The data was obtained from the central waste bank in Makassar City, covering daily purchases of waste from a number of waste bank units, which were summarized based on waste type codes for the period January 2023 to June 2025. Of the 180 waste bank units recorded, there was an average of 37 waste bank units that sold waste types such as plastic, paper, glass bottles, metal, and used cooking oil. The data items from the waste purchase transaction data include: waste collection date, waste bank unit, waste type coded based on its characteristics, volume of waste collected (kg), and total price. The recapitulation of waste purchases is recorded in the table format shown in Table 1.

Table 1. Recapitulation of inorganic waste purchases

No	Weighing date	Name of Waste Bank Unit/Neighborhood/RW	Inorganic waste product code (plastic)													
			P01B		P01K		P02B		P02C		P03B		P04B		P04C	
			Kg	IDR	Kg	IDR	Kg	IDR	Kg	IDR	Kg	IDR	Kg	IDR	Kg	IDR
1	January 2, 2023	CITRA MAPALA	14	84,000	5	16,000	-		1	1,900	3	18,300	3	13,800	...	...
2		SMART	2	12,000	-		1	2,000	1	1,900	12	73,200	4	18,400	...	...
3		MAPALA IMAGE	24	144,000	-		3	6,000	-		16	97,600	29	133,400	...	...
4		MAKARTI	19	114,000	-		-		-		41	250,100	29	133,400	...	...
5		MAKARTI	17	102,000	-		20	40,000	3	5,700	45	274,500	34	156,400	...	...

4. RESULTS AND DISCUSSION

4.1. Dataset description

The dataset used in this study consists of waste bank transaction data comprising 45,544 records with three main attributes, namely codes, quantity, and value. The dataset covers 44 unique types of waste, representing the variety of materials managed in the waste bank system, as shown in Table 2. The quantity attribute has 495 value variations, while the value attribute has a fairly wide range, from IDR 200 to IDR 4,815,000. This variation shows the heterogeneity of transaction activities and potential differences in economic value between types of waste.

Figure 2 shows a trend of waste purchase values that fluctuate significantly across all observations. Most transactions are in the low to medium economic value range, while there are several sharp spikes indicating purchases of waste with very high economic value. This pattern indicates that the distribution of the economic value of waste is uneven and tends to be dominated by low-value waste, while high-value waste is relatively scarce.

Table 2. Waste bank transaction data

No	Codes	Quantity	Value
1	B01	1	1,100
2	B01	2	2,200
3	B01	4	4,400
4	B01	8	8,800
5	B01	159	174,900
...	...	...	...
45541	P20	3	5,700
45542	P20	6	11,400
45543	P20	5	9,500
45544	P20	8	15,200
45545	P20	3	5,700

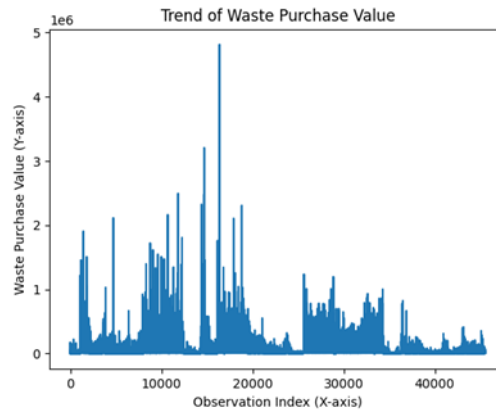


Figure 2. Waste purchase trends

4.2. Descriptive statistical analysis

Descriptive statistical analysis was performed to understand the characteristics of the data distribution before modeling. A summary of the transaction value statistics is shown in Table 3. A higher average value than the median indicates an asymmetrical distribution. A relatively large standard deviation (compared to the median) indicates significant variation in transaction values.

Table 3. Descriptive statistics of transaction values

Parameter	Value
Number of data	45,544
Maximum	200
First quartile (Q1)	6,000
Median	17,500
Third quartile (Q3)	54,600
Maximum	4,815,000
Average	51,693

4.3. Transaction value distribution

Based on the histogram of transaction value distribution in Figure 3, it can be seen that the data has a right-skewed pattern. Most transactions are concentrated in the low value range, while the frequency of transactions decreases significantly as the transaction value increases. This shows that waste bank activities are dominated by small to medium value transactions. The distribution also shows a small number of transactions with very high values that form a long tail on the right side of the graph. The existence of these extreme values causes the data range to be very wide and indicates the presence of outliers, as shown in Figure 4. This condition explains why the mean value tends to be higher than the median and reinforces the indication that the distribution is not normal.

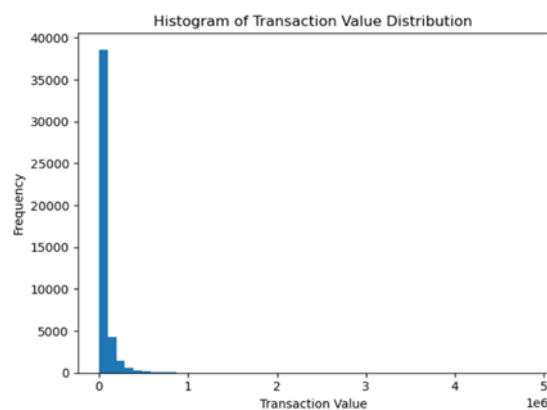


Figure 3. Distribution of transaction values

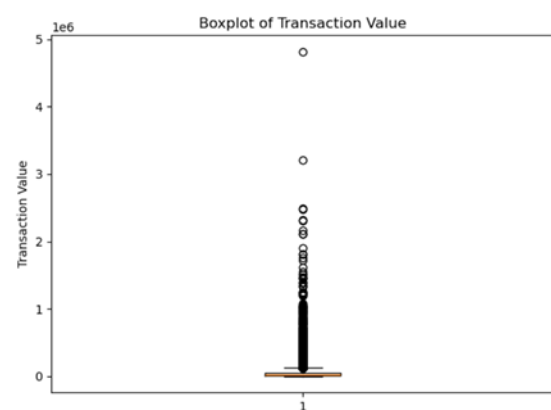


Figure 4. Boxplot of transaction value

In practical terms, this distribution pattern reflects the economic characteristics of waste banks, where most deposits involve small volumes and low transaction values. Conversely, high-value transactions only occur in specific cases, such as large deposits or waste types with high selling prices. This asymmetric characteristic is a crucial consideration in modeling, making decision tree-based algorithms like random forest and gradient boosting more suitable for handling abnormal distributions and outliers.

4.4. Pre-processing

Before labeling and modeling, the dataset goes through several pre-processing stages to ensure data quality and consistency. These stages include data cleaning, data type transformation, categorical variable coding, and dataset division. The initial pre-processing stage is carried out to identify and address data quality issues, including:

- i) Missing value check: one missing value was found in the value attribute. This data was deleted because the amount was very small ($< 0.01\%$) and did not affect the overall distribution of the dataset.
- ii) Data type validation: the quantity attribute was initially non-numeric (object), so it was converted to a numeric type to ensure compatibility with the machine learning algorithm.
- iii) Outlier check: boxplot analysis shows extreme values in the value attribute. However, outliers are not removed because these values represent high-value real transactions and are still relevant in the context of the circular economy.
- iv) Categorical variable encoding: the codes attribute is converted into a numerical representation using a technique so that it can be processed by a decision tree-based algorithm.

4.5. Data labeling

Since the waste bank transaction dataset does not have explicit economic value category labels, this study applies two labeling strategies to build target classification categories, namely:

- i) Clustering-based labeling (k-means): in the first strategy, the k-means algorithm was applied to the main numerical attributes, namely quantity and value. The number of clusters was set at $k = 3$, assuming that the data could be grouped into three economic value categories (low, medium, and high). The clustering results produced three centroids with significantly different average value characteristics. After the clustering process, each cluster is mapped into an economic category based on the average value in each cluster, namely: cluster with the lowest average $\rightarrow$ low, cluster with the medium average $\rightarrow$ medium, and cluster with the highest average $\rightarrow$ high.
- ii) Quantile-based classification: the second strategy uses a statistical distribution-based approach to the value attribute. Category boundaries are determined based on the distribution quartiles as follows: low: $\text{value} \leq Q1$, medium: $Q1 < \text{value} \leq Q3$, high: $\text{value} > Q3$.

The comparison results presented in Table 4 show that k-means-based labeling produces a highly unbalanced class distribution, where 87.25% of the data is classified in the low category and only 0.84% in the high category. This condition has the potential to cause model bias and reduce classification ability in minority classes. In contrast, the value distribution-based approach produces a relatively proportional class distribution (low 25.48%, medium 49.52%, and high 25.00%), thereby supporting the stability of model training and evaluation. Therefore, the quantile-based labeling strategy is used in the implementation of the random forest and gradient boosting models in this study.

Table 4. Comparison of label distributions

Strategy	Low	Moderate	High
K-means	39,733	5,423	384
Quantile-Based	11,604	22,552	11,384

4.6. Data split

The dataset is divided into 70% training data and 30% test data. The division is done using the stratified train-test split method, so that the proportions of low, medium, and high classes in the training data and test data remain consistent. This strategy aims to avoid class distribution bias and improve model evaluation reliability.

4.7. Modeling with random forest and gradient boosting algorithms

4.7.1. Model parameter configuration

In this study, both ensemble algorithms were configured using key parameters aimed at optimizing the balance between model complexity and generalization ability. Based on the parameter configuration shown in Table 5, there is a difference in the philosophy of the two methods. Random forest uses deeper trees

(depth 10) with a parallel approach (bagging). Gradient boosting uses shallower trees (depth 5) but is built sequentially with a controlled learning rate. The combination of 200 trees in both models ensures sufficient representation capacity to capture the distribution patterns of waste economic values, while still being controlled to prevent overfitting.

Table 5. Random forest and gradient boosting model parameter configuration

Model	Parameter	Value	Description
Random forest	n_estimators	200	Number of trees in the ensemble to improve stability and reduce variance
	max_depth	10	Maximum depth of each tree to control model complexity
	min_samples_leaf	10	Minimum number of samples in each leaf to prevent overfitting
	class_weight	Balanced	Class weight adjustment to address potential data imbalance
Gradient boosting	random_state	42	Ensuring experiment reproducibility
	n_estimators	200	Number of trees built sequentially
	learning_rate	0.1	Learning rate to control the contribution of each tree
	max_depth	5	Maximum depth of trees to keep the model generalizable
	random_state	42	Ensures reproducibility of experiments

4.7.2. Formation of one of the trees

The formation of one of the decision trees in the random forest model is presented in Figure 5, where the hierarchical structure utilizes the Quantity_scaled feature as the main separator at the root node, followed by Codes_enc in the next branch. This pattern indicates that the volume/amount of waste is the dominant determinant in economic value classification, while the type of waste acts as a purifying factor at a deeper level. The class distribution at the leaf node shows a relatively clear separation between the low, medium, and high categories, with the majority of terminal nodes dominated by a particular class. The fairly complex depth of the tree demonstrates the model’s ability to capture non-linear patterns and interactions between features. Overall, this tree structure is consistent with the results of the feature importance analysis, which confirms that transaction quantity has the most significant contribution in determining the economic value category of waste.

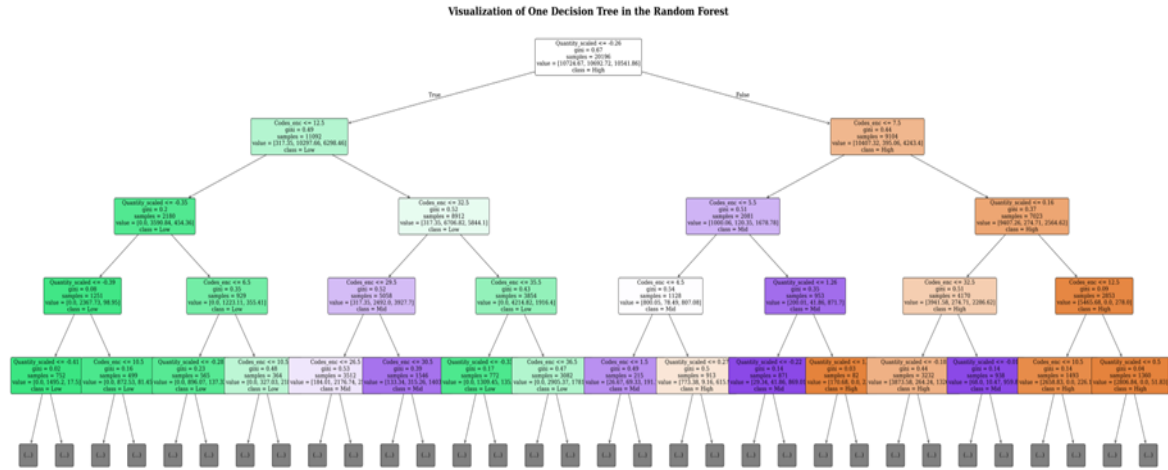


Figure 5. Visualization of one decision tree in the random forest model

The visualization of one of the decision trees in the gradient boosting model in Figure 6 shows a relatively controlled tree structure with initial separation dominated by the Quantity_scaled feature, which is then refined by Codes_enc at the next branching level. Unlike random forest, which builds many independent trees, trees in gradient boosting are formed sequentially to correct prediction errors from previous iterations, so that each branching represents a residual correction process. The consistent split pattern in transaction quantity confirms that waste volume is a major factor in determining economic value categories, while waste type serves as a classification adjustment at the distribution boundary conditions. The class distribution at the terminal node shows a fairly clear separation between categories, reflecting the model’s ability to gradually capture non-linear patterns through an adaptive and targeted boosting process.

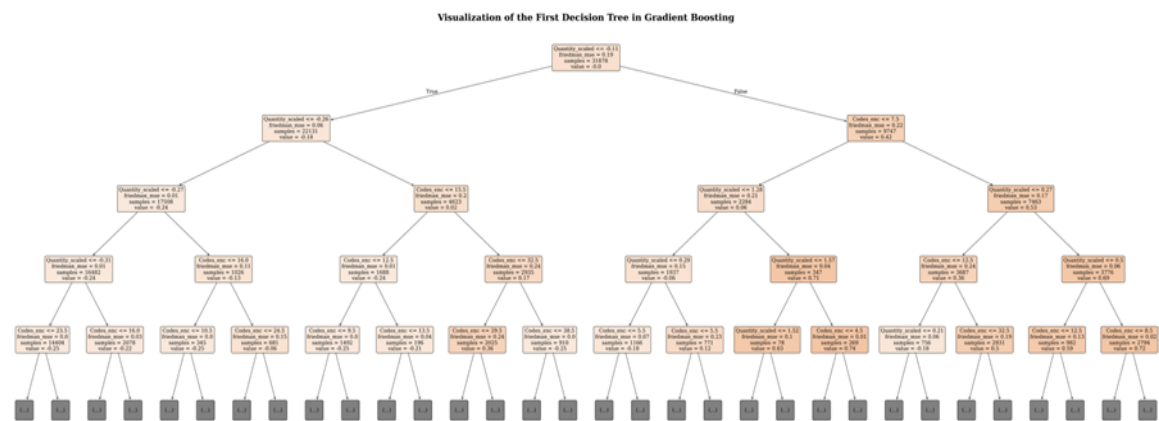


Figure 6. Visualization of one decision tree in the gradient boosting model

4.7.3. Model evaluation results

The evaluation results in Figure 7 show that both ensemble models, random forest and gradient boosting, are able to classify the economic value of waste into low, medium, and high categories with excellent performance. However, gradient boosting consistently shows superior e performance compared to random forest. An accuracy improvement of ± 0.8% indicates that the boosting approach has better generalization capabilities than the bagging approach in the research dataset.

	precision	recall	f1-score	support	Model
High	0.913721	0.961347	0.936929	3415.000000	Random Forest
Low	0.881316	0.962080	0.919929	3481.000000	Random Forest
Mid	0.957888	0.887526	0.921366	6766.000000	Random Forest
accuracy	0.924974	0.924974	0.924974	0.924974	Random Forest
macro avg	0.917642	0.936984	0.926074	13662.000000	Random Forest
weighted avg	0.927338	0.924974	0.924890	13662.000000	Random Forest
High	0.959153	0.942020	0.950510	3415.000000	Gradient Boosting
Low	0.918601	0.943407	0.930839	3481.000000	Gradient Boosting
Mid	0.941334	0.936743	0.939033	6766.000000	Gradient Boosting
accuracy	0.939760	0.939760	0.939760	0.939760	Gradient Boosting
macro avg	0.939696	0.940723	0.940127	13662.000000	Gradient Boosting
weighted avg	0.939996	0.939760	0.939814	13662.000000	Gradient Boosting

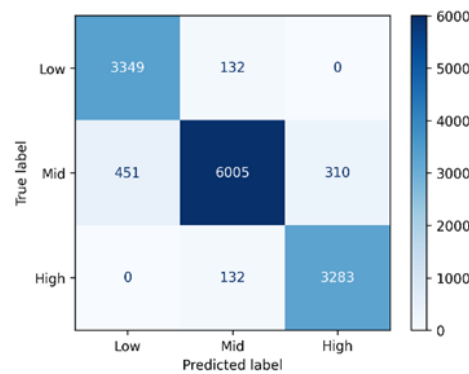
Figure 7. Comparison of metrics

The results of the analysis per class as presented in Figure 8 show that both models are capable of classifying economic value categories with balanced performance, but gradient boosting consistently provides higher results across all classes. In the low class, the difference in performance is relatively small, although gradient boosting has a slightly better F1-score, indicating a more optimal balance between precision and recall. In the medium class, the most significant improvement is seen, where gradient boosting is able to significantly increase the F1-score compared to random forest, indicating its superior ability to handle classes with overlapping distributions. Meanwhile, in the high class, both models show very strong performance, but gradient boosting remains superior in precision and F1-score, meaning that this model is more accurate in identifying high economic value waste. Overall, these results confirm that the boosting approach is more effective in distinguishing adjacent classes and handling the complexity of data distribution.

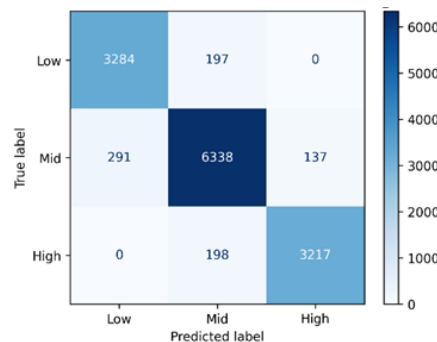
	Codes_enc	Quantity_scaled	Actual_Label	Pred_RF	Pred_GB
37217	36	0.432466	High	High	High
6875	6	-0.325245	Low	Low	Low
37562	36	-0.382073	Mid	Mid	Mid
33868	32	-0.363130	Mid	Mid	Mid
6927	6	-0.306302	Low	Low	Low

Figure 8. Results of analysis by class

Comparison of the confusion matrices in Figure 9 shows that both models classify the three classes well, although with slightly different error patterns. Random forest performs more consistently in the low and medium classes, while gradient boosting achieves more accurate predictions in the high class. Overall, random forest is more conservative and stable, while gradient boosting is more effective in identifying high-value waste. Figure 9(a) shows the confusion matrix for random forest. The model correctly classified 3,283 low class transactions and 3,349 mid class transactions, with most errors occurring near the boundary with the high class. This confirms that random forest produces stable and conservative predictions. Figure 9(b) shows the confusion matrix for gradient boosting. The model correctly classified 3,217 low class transactions, 3,284 mid class transactions, and 6,338 high class transactions. Although showing slightly more errors in the low and medium classes, the model performs better in the high class, showing stronger adaptability at the class boundaries.



(a)



(b)

Figure 9. Confusion matrix evaluation for (a) random forest and (b) gradient boosting

Feature importance analysis based on Figure 10 shows a consistent pattern in both models. In random forest, the contribution of Quantity_scaled is 0.63 and Codes_enc is 0.37. In gradient boosting, the contribution of Quantity_scaled is around 0.60 and Codes_enc is around 0.40. These results indicate that transaction quantity is the main determinant in determining the economic value category of waste, while waste type acts as a supporting factor that refines the classification. These findings reinforce that the economic value of waste is more influenced by transaction volume than by material type alone.

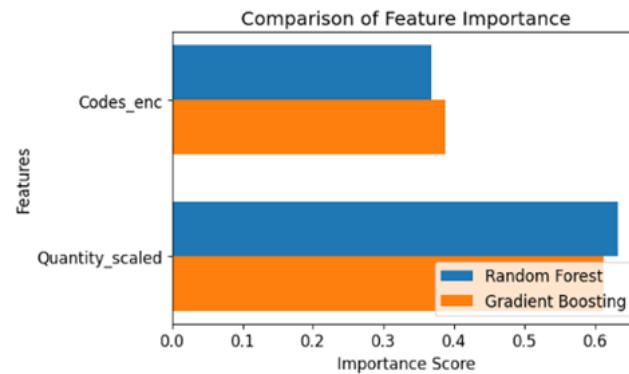


Figure 10. Analysis results by class

4.8. Practical implications

Random forest and gradient boosting both classified waste economic value with excellent accuracy across all three categories, confirming that quantity is a strong predictor of purchase value. Random forest gave more stable, consistent predictions due to its ensemble-of-trees mechanism, consistent with prior studies on waste classification [39]. Gradient boosting, meanwhile, captured more complex patterns and handled varied data better through its iterative error-correction mechanism [40]. The quantile-based class distribution helped both models learn effectively, though random forest generalized more stably while gradient boosting was more sensitive to data variation and detail.

Practically, these models can help waste bank managers and local governments quickly and systematically predict waste economic value, supporting sorting, recycling planning, and data-driven decisions that strengthen circular economy efforts at the city level. The study has some limitations: it relies on secondary data of variable quality, and only uses purchase quantity and value, without external factors like market prices or geography. Future research could incorporate spatial data, market trends, and other external variables, and develop an application-based system to support practical implementation of these prediction models.

5. CONCLUSION

This study developed an economic value prediction model for waste based on machine learning using random forest and gradient boosting with category labeling through k-means clustering. The evaluation results showed that both models performed very well with an accuracy of over 92%. Gradient boosting consistently showed superior performance, especially in improving the F1-score and reducing classification errors in high-value waste categories. Confusion matrix analysis shows that random forest is more stable, while gradient boosting is more adaptive to the distribution boundaries between classes. In addition, transaction quantity proved to be a dominant factor in determining economic value categories. These findings confirm that the ensemble learning approach is effective in supporting data-driven decision-making systems in waste bank management and strengthening the circular economy at the local level.

ACKNOWLEDGEMENTS

The author would like to express his deepest appreciation and gratitude to the Directorate of Research and Community Service of the Ministry of Higher Education, Science, and Technology for its financial support, which enabled the implementation of this research. The author would also like to thank the Makassar City Environment Agency, particularly the managers of the Makassar City Waste Bank, for their support in providing data, cooperation, and facilitation during the data collection and validation process. The support and collaboration from various parties greatly contributed to the smooth running and success of this research.

FUNDING INFORMATION

This research was funded by Directorate of Research and Community Service of the Ministry of Higher Education, Science, and Technology through the Fundamental Research Grant scheme (decree number 130/C3/DT.05.00/PL/2025 and contract number 0034/UNDIPA/G1/VI/2025).

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Role Taxonomy (CRediT) to acknowledge the contributions of each author, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Andi Irmayana	✓	✓	✓			✓		✓	✓	✓			✓	✓
Nurul Aini	✓	✓		✓	✓			✓	✓	✓				
Ramlah P.	✓			✓	✓		✓			✓		✓		
Muhammad Husnul		✓	✓		✓	✓		✓		✓	✓		✓	✓
Yaqin														
Salsabila Sabir	✓	✓		✓	✓		✓		✓		✓	✓		✓

C : Conceptualization	I : Investigation	Vi : Visualization
M : Methodology	R : Resources	Su : Supervision
So : Software	D : Data Curation	P : Project administration
Va : Validation	O : Writing - Original Draft	Fu : Funding acquisition
Fo : Formal analysis	E : Writing - Review & Editing	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author, [AI].

REFERENCES

[1] M. F. Muhlis, M. Selinrung, and S. Syafri, "Partisipasi masyarakat terhadap pengelolaan sampah di Kota Makassar," *Urban and Regional Studies Journal*, vol. 4, no. 2, pp. 110–115, 2022, doi: 10.35965/ursj.v4i2.1466.

[2] M. Kumar *et al.*, "A review on value-addition to plastic waste towards achieving a circular economy," *Science of the Total Environment*, vol. 921, 2024, doi: 10.1016/j.scitotenv.2024.171106.

[3] X. Jiang and B. Bateer, "A systematic review of plastic recycling: technology, environmental impact and economic evaluation," *Waste Management and Research*, vol. 43, no. 8, pp. 1159–1178, 2025, doi: 10.1177/0734242X241310658.

[4] R. Gatta *et al.*, "The transformation of roles and capacities of housewives in household waste management in Makassar City," *Extension Journal*, vol. 18, no. 2, pp. 265–276, 2022.

[5] H. Luo, H. Tyrrell, J. Bai, R. I. Muazu, and X. Long, "Fundamental, technical and environmental overviews of plastic chemical recycling," *Green Chemistry*, vol. 26, no. 23, pp. 11444–11467, 2024, doi: 10.1039/d4gc03127j.

[6] F. Niu *et al.*, "State-of-the-art and perspectives of hydrogen generation from waste plastics," *Chemical Society Reviews*, vol. 54, no. 10, pp. 4948–4972, 2025, doi: 10.1039/d4cs00604f.

[7] E. Desnia, E. Rosie, S. B. Hartono, W. F. Simanullang, A. A. Anggorowati, and S. Lourentius, "Optimization of pyrolysis of polypropylene and polyethylene based plastic waste become an alternative oil fuel using bentonite catalyst," *E3S Web of Conferences*, vol. 475, 2024, doi: 10.1051/e3sconf/202447505006.

[8] M. V. Singh, B. Devi, Z. Wu, C. Sharma, H. Liu, and D. Pan, "Transforming waste plastic into high-value petrochemical and diesel fraction through the pyrolysis process," *ES Energy and Environment*, vol. 26, pp. 1–22, 2024, doi: 10.30919/esee1208.

[9] T. Imran, N. Nurhaedar, A. Syam, M. Dahri, and M. Noor, "Barang rongsokan dan peluang bisnis berbasis ekonomi sirkular di Sulawesi Selatan: tinjauan literatur," *Al-Hubuts*, vol. 21, no. 1, pp. 266–283, 2025, doi: 10.30603/ab.v21i1.6468.

[10] T. R. Ling, Y. Y. Li, C. M. Tsai, and T. T. Hung, "Biodiesel produced from waste cooking oil by microwave transesterification using inert aluminium foil (1%) as a heating promoter," *Sustainable Chemistry and Pharmacy*, vol. 41, 2024, doi: 10.1016/j.scp.2024.101738.

[11] A. Ahuja, P. Samyn, and V. K. Rastogi, "Paper bottles: potential to replace conventional packaging for liquid products," *Biomass Conversion and Biorefinery*, vol. 14, no. 13, pp. 13779–13805, 2024, doi: 10.1007/s13399-022-03642-3.

[12] J. Zhang, G. Zhang, M. Bhuiyan, S. Wang, and C. Q. Li, "Temporal fluctuations in waste generation: an environmental and cost evaluation of glass waste collection and recycling in glass containers and asphalt production," *Waste Management*, vol. 206, 2025, doi: 10.1016/j.wasman.2025.115040.

[13] H. Putra and R. Rumini, "Comparative study of logistic regression, random forest, and XGBoost for bank loan approval classification," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2822–2835, Oct. 2025, doi: 10.30871/jaic.v9i5.10862.

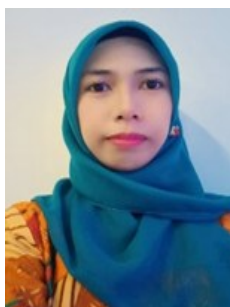
[14] B. Mohammed and C. Hamza, "A robust estimation of blasting-induced flyrock using machine learning decision tree algorithms: random forest, gradient boosting machine, and XGBoost," *Mining, Metallurgy and Exploration*, vol. 42, no. 3, pp. 1609–1624, 2025, doi: 10.1007/s42461-025-01267-4.

[15] C. Rudin, "Leo Breiman, the rashomon effect, and the Occam Dilemma," 2025, *arXiv: 2507.03884*.

[16] A. Waladi, "Peningkatan akurasi klasifikasi tutupan lahan menggunakan random forest pada data Sentinel-2 di Jambi," *Jurnal Fasilkom*, vol. 15, no. 1, pp. 17–24, 2025, doi: 10.37859/jf.v15i1.8886.

- [17] D. Chutia, N. Borah, D. Baruah, D. K. Bhattacharyya, P. L. N. Raju, and K. K. Sarma, "An effective approach for improving the accuracy of a random forest classifier in the classification of hyperion data," *Applied Geomatics*, vol. 12, no. 1, pp. 95–105, 2020, doi: 10.1007/s12518-019-00281-8.
- [18] A. O. Alqabbany and A. M. Azmi, "Measuring the effectiveness of adaptive random forest for handling concept drift in big data streams," *Entropy*, vol. 23, no. 7, pp. 1–24, 2021, doi: 10.3390/e23070859.
- [19] Y. A. Saadoun and R. H. Abdulmir, "Improved random forest algorithm performance for big data," *Journal of Physics: Conference Series*, vol. 1897, no. 1, 2021, doi: 10.1088/1742-6596/1897/1/012071.
- [20] A. Labovich, "A case for library-level k-means binning in histogram gradient-boosted trees," *Transactions on Machine Learning Research*, 2025, pp. 1-20.
- [21] T. Champahom *et al.*, "Deep learning vs. gradient boosting: optimizing transport energy forecasts in Thailand through LSTM and XGBoost," *Energies*, vol. 18, no. 7, pp. 1–30, 2025, doi: 10.3390/en18071685.
- [22] N. Rasif, W. Utama, S. A. Garini, R. F. Indriani, and D. P. N. Putra, "Prediction of sonic log values using a gradient boosting algorithm in the 'AB' field," *BIO Web of Conferences*, vol. 157, 2025, doi: 10.1051/bioconf/202515707002.
- [23] I. Ranjbar, Y. Ventikos, and M. Arashpour, "Deep learning-based construction and demolition plastic waste classification by resin type using RGB images," *Resources, Conservation and Recycling*, vol. 212, 2025, doi: 10.1016/j.resconrec.2024.107937.
- [24] D. V. Yevle and P. S. Mann, "Artificial intelligence-based waste management: a review of classification, techniques, issues, and challenges," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 15, no. 2, 2025, doi: 10.1002/widm.70025.
- [25] M. Nahiduzzaman *et al.*, "An automated waste classification system using deep learning techniques: toward efficient waste recycling and environmental sustainability," *Knowledge-Based Systems*, vol. 310, 2025, doi: 10.1016/j.knsys.2025.113028.
- [26] I. Dawar, A. Srivastava, M. Singal, N. Dhyani, and S. Rastogi, "A systematic literature review on municipal solid waste management using machine learning and deep learning," *Artificial Intelligence Review*, vol. 58, 2025, doi: 10.1007/s10462-025-11196-9.
- [27] A. H. Hidayah, D. Novitasari, R. Kamila, Trimono, and M. Nasrudin, "Cluster modeling with k-means on provincial data in Indonesia based on environmental indicators," *Journal of Artificial Intelligence and Engineering Applications*, vol. 4, no. 3, pp. 2249–2245, 2025, doi: 10.59934/jaiea.v4i3.1143.
- [28] R. Magriaty, K. Murtalaksono, and S. Anwar, "Analisis k-means cluster untuk identifikasi kawasan pengelolaan sampah di Kabupaten Tapin Provinsi Kalimantan Selatan," *Journal of Regional and Rural Development Planning*, vol. 7, no. 1, pp. 79–90, 2023, doi: 10.29244/jp2wd.2023.7.1.79-90.
- [29] T. D. Khuc, N. D. Luong, D. H. Dang, and V. A. Tran, "Comparison of random forest and extreme gradient boosting algorithms in land cover classification in Van Yen district, Yen Bai province, Vietnam," *Journal of Hydro-meteorology*, vol. 6, no. 23, pp. 50–59, 2025, doi: 10.36335/vnjhm.2025(23).50-59.
- [30] M. V. Lau, A. D. Handayani, M. Y. Widodo, R. I. Handayani, and R. L. Pratiwi, "Implementasi metode random forest untuk klasifikasi sampah organik dan anorganik berbasis citra digital," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 4, pp. 400–406, 2025, doi: 10.31004/riggs.v4i4.3410.
- [31] N. Maulidah, M. Maulidah, R. Supriyadi, H. Nalatissifa, S. Diantika, and A. Fauzi, "Prediksi kualitas air menggunakan metode random forest, decision tree, dan gradient boosting," *Jurnal Khatulistiwa Informatika*, vol. 12, no. 1, pp. 1–6, 2024, doi: 10.31294/jki.v12i1.16004.
- [32] X. C. Nguyen *et al.*, "Development of machine learning-based models to forecast solid waste generation in residential areas: a case study from Vietnam," *Resources, Conservation and Recycling*, vol. 167, 2021, doi: 10.1016/j.resconrec.2020.105381.
- [33] H. A. Salman, A. Kalakech, and A. Steiti, "Random forest algorithm overview," *Babylonian Journal of Machine Learning*, pp. 69–79, 2024, doi: 10.58496/BJML/2024/007.
- [34] R. Nugraha, N. Suarna, I. Ali, and D. Rohman, "Optimasi pengelolaan sampah melalui model pengelompokan dengan algoritma k-means," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, 2025, doi: 10.23960/jitet.v13i1.5694.
- [35] E. M. Ahmed, K. E. Elfaki, and E. Abusham, "Machine learning models predictive performance of Asian economies' green technological progress," *Sustainable Futures*, vol. 10, 2025, doi: 10.1016/j.sfr.2025.101323.
- [36] A. Geron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*, 3 ed, Sebastopol, California: O'Reilly Media, Inc., 2022.
- [37] S. Prusty, S. Patnaik, and S. K. Dash, "SKCV: Stratified k-fold cross-validation on ML classifiers for predicting cervical cancer," *Frontiers*, vol. 4, pp. 1–12, 2022, doi: 10.3389/fnano.2022.972421.
- [38] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing and Management*, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.
- [39] A. M. Akbar, M. Basri, and Wahyuddin, "Implementasi machine learning menggunakan algoritma klasifikasi untuk mendeteksi jenis sampah," *Jurnal Publikasi Manajemen Informatika*, vol. 3, no. 3, 2025, doi: 10.55606/jupumi.v3i3.3751.
- [40] A. Angdresey, I. Y. Kairupan, and A. G. Mongkareng, "Leveraging convolutional neural networks for multiclass waste classification," *Journal of Applied Informatics and Computing*, vol. 9, no. 4, pp. 1137–1145, 2025, doi: 10.30871/jaic.v9i4.9373.

BIOGRAPHIES OF AUTHORS



Andi Irmayana    completed a master's degree in Electrical Engineering with a concentration in Information Technology in 2011 at Universitas Hasanuddin, Makassar. Previously, she completed her bachelor's degree in 2008 in the Department of Information Systems, Universitas DIPA Makassar. She actively teaches courses in programming algorithms, database systems, decision support systems, and professional programming. Her research interests in science and technology include system design, big data science, and machine learning. Active as a member of the Association of Informatics and Computer Colleges and the Indonesian Computer, Electronics, and Instrumentation Support Society (INDOCEISS). She can be contacted at email: irmayana.andi@undipa.ac.id.



Nurul Aini    interest in computer science began in 2005. This led the author to pursue higher education at Universitas DIPA Makassar, majoring in Information Systems, and successfully graduating in 2009. She then continued postgraduate studies in the Master's Program of Informatics Engineering at Universitas Hasanuddin and completed the program in 2013. She is also actively involved in teaching system analysis and design, object-oriented programming, and startup project courses in the Department of Software Engineering. She can be contacted at email: nurulaini.m11@undipa.ac.id.



Ramlah P.    was registered as a graduate of the postgraduate program in Financial Management at Universitas Hasanuddin Makassar and in 2005, she was registered as a graduate of the bachelor's program at STIEM Bongaya Makassar, Management Study Program. Since 2008, she has been a permanent lecturer at the foundation and is currently placed in the Department of Digital Business, Universitas Teknologi Akba Makassar. She actively teaches the courses introduction to management, human resource management, introduction to economics, accounting, financial management. Besides being a permanent lecturer at the foundation, she also active as a member of the Kartini Ujung Pandang Education Foundation from 2008 until now. She can be contacted at email: ramlah@unitama.ac.id.



Muhammad Husnul Yaqin    is currently pursuing a bachelor's degree in Information Systems at Universitas DIPA Makassar. He previously completed his secondary education at SMA Negeri 4 Makassar. During his high school years, he was actively involved as the head of the Information and Computer Technology (ICT) extracurricular program from 2021 to 2023. His academic interests focus on information systems, programming, and technology development. He is enthusiastic about exploring the fields of web development, database systems, and artificial intelligence. He can be contacted at email: husnulyaqin1936@gmail.com.



Zalsabila Sabir    is an active undergraduate student in the Department of Information Systems, Universitas DIPA Makassar and is currently in her 5th semester. She previously completed her vocational high school education at SMKN 04 Makassar majoring in accounting, where she also carried out an internship at the Human Resources Development Agency (BPSDM). Her academic interests include database, web programming, and decision support systems. During her studies, she has participated in various academic activities focusing on the development of analytical and programming skills. She can be contacted at email: zalsabilas9@gmail.com.