PREDICTION OF PLASTIC-TYPE FOR SORTING SYSTEM USING DECISION TREE MODEL

ASTUTI[1], ANTHONY COSTA[2], AKBAR TEGUH PRAKOSO[1], IRSYADI YANI[1], YULIA RESTI[3,*]

[1]Department of Mechanical Engineering, Faculty of Engineering, Universitas Sriwijaya, Sumatera Selatan, Indonesia

[2]Department of Civil Engineering, Faculty of Engineering, Universitas Sriwijaya, Sumatera Selatan, Indonesia

[3]Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Sriwijaya, Sumatera Selatan, Indonesia

**Corresponding author: yulia_resti@mipa.unsri.ac.id*

(Received: 3 January 2023; Accepted: 27 February 2023; Published on-line: 1 March 2023)

ABSTRACT: Plastic is the most widely used inorganic material globally, but its hundred-year disintegration time can harm the environment. Polyethylene Terephthalate (PET/PETE), High-Density Polyethylene (HDPE), and Polypropylene are all commonly used plastics that have the potential to become waste (PP). An essential first step in the recycling process is sorting out plastic waste. A low-cost automated plastic sorting system can be developed by using digital image data in the red, green, and blue (RGB) color space as the dataset and predicting the type using learning datasets. This paper proposes the Decision Tree model to predict the three plastic-type sorting systems based on discretizing predictor variables into two and three categories. The resampling method of k-fold cross-validation with ten folds for less biased. Discretization of the predictor variables into three categories informs that the proposed decision tree model has higher performance compared to the two categories with an accuracy of 81.93 %, a recall-micro of 72.89 %, a recall-macro of 72.30 %, a specificity-micro of 86.45%, and the specificity-macro of 86.51%, respectively. The micro is determined by the number of decisions made for each object. In comparison, the macro is calculated based on the average decision made by each class.

KEY WORDS: Decision Tree, Discretization, Plastic-Type, Prediction

1. INTRODUCTION

Plastic is the most widely used inorganic material globally, and its usage has been increasing rapidly, particularly in countries experiencing rapid economic growth [1]. This is because plastic is versatile, lightweight, and cost-effective, making it an attractive material for various industries, including packaging, construction, and automotive [2]. However, the rapid increase in plastic usage has also resulted in a significant environmental problem, as plastic waste is not biodegradable and can persist in the environment for hundreds of years [3]. Plastic waste threatens wildlife, marine ecosystems, and human health, and its accumulation in landfills and oceans has become a global concern [4]. Various measures have been taken to address this problem, including plastic recycling and the development of sorting systems. Sorting systems are designed to separate plastic waste into different types based on their properties, such as color, texture, and shape [5]. This enables plastic waste to be recycled more efficiently and effectively. This stage is critical because the improper classification of plastic types can result in cross-contamination, increasing industrial operating costs. In addition, this

https://doi.org/10.51630/ijes.v4i1.86

process frequently encounters difficulties when attempting to differentiate between different types of plastic. The plastic types Polyethylene Terephthalate (PET/PETE), High-Density Polyethylene (HDPE), and Polypropylene (PP) are widely used in the community and have the potential to become waste [6].

Automatic plastic sorting is a viable solution to the ineffectiveness and inefficiency of the manual method using human power. A low-cost automatic plastic sorting system can be developed by utilizing machine learning and a digital image with the Red, Green, and Blue (RGB) color model as a dataset. Many researchers are developing sorting systems for plastic waste to improve the efficiency and effectiveness of plastic waste recycling. Machine learning-derived predicted plastic-type values have a purpose in the sorting process. Khona'ah et al. (2019) and Yani et al. (2020) developed artificial neural network backpropagation (ANNB) method to predict plastic type based on digital images [7][8]. The ANNB algorithm is a widely used and popular prediction/classification algorithm. However, the minimum accuracy of the classification method is 85 % [9].

Additionally, the performance of the method is solely based on its accuracy. Therefore, numerous metrics must be used to evaluate the effectiveness of methods. One of the prediction methods in machine learning is decision tree model analysis. This task predicts the categorical target variable [10]. Wang et., al (2019) propose using decision tree learning to improve the accuracy of these models to predict the duration of leaf wetness in the greenhouse environment [11]. The results show that the decision tree model can significantly improve the accuracy of models that predict leaf wetness duration.

Furthermore, the decision tree is a fast learning speed method and requires little or no data preprocessing [12]. This paper proposes the Decision Tree model to predict the three plastic-type sorting systems based on discretizing predictor variables into two and three categories. A resampling method of k-fold cross-validation with ten folds for less biased [13].

2. EXPERIMENTAL SECTION

2.1. Materials and Method

Four hundred fifty plastic data were collected by capturing the images in three random poses. Plastic waste comes from three types; PET, HDPE, and PP. The material properties of plastic material as tabulated in Table 1. The image of each plastic type is given in Fig. 1, and the statistics summary of image data collected related to the five normalized predictor variables is noted in Table 1. The code includes XX_{11}, XX_{22}, XX_{33}, XX_{44}, and XX_{55}, which represent standardized feature variables of the average pixel of red, green, blue, entropy, and variance, respectively.

(a) (b) (c)

Fig. 1. Digital images of plastic-type: (a) Polyethylene Terephthalate, (b) high-density polyethylene, and (c) polypropylene

Table 1: Summary statistic of variable

Statistic	Predictor Variable				
	Red (XX_{11})	**Green (XX_{22})**	**Blue (XX_{33})**	**Entropy (XX_{44})**	**Variance (XX_{55})**
Minimum	0.33	0.35	0.33	0.00	0.00
1st Quartile	0.61	0.63	0.67	0.01	0.01
Median	0.70	0.75	0.78	0.02	0.02
Mean	0.76	0.79	0.80	0.01	0.05
3rd Quartile	0.98	0.99	0.98	0.02	0.12
Maximum	1.00	1.00	1.00	0.03	0.13

The obtained images are processed into RGB color format, where each color component has a value of 8 bits so that each color component has a scale of $2^8 = 256$ or a pixel value range of 0 to 255. The image's resolution stored in the database is 560 × 420 pixels. The image is cropped to 34 × 34 pixels with cropping coordinates [280 180 33 33]. Fig. 2 presents the three types of cropped plastic waste digital images. To predict the plastic-type using Decision Tree, we follow the steps. First, determine entropy using (a). Suppose where S and Sc be the total number of plastic images and the total number of plastic images in the c-th split of the predictor variable X. Let P_j, and P_c be the prior probability in the j-th type of the predictor variable X, and the prior probability in the c-th split of the predictor variable X, respectively. Entropy S and S_c are each formulated as [14]:

$$Entropy\ (S) = \sum_{j=1}^{K_s} - P_j log_2 P_j \tag{1}$$

Second, determine information gain using (2),

$$Information\ Gain\ (S,X) = Entropy\ (S) = \sum_{c=1}^{K_x} \frac{|S_c|}{|S|} Entropy\ (S_c) \tag{2}$$

Third, choose the variable that has the largest Information Gain value. Next, form a node that contains the variable. The performance measurements are used to represent classification performance in various metrics such as accuracy, recall-micro (μ), recall-macro (M), specificity-micro (μ), and specificity-macro (M). The TP_j, FP_j, TN_j, and FN_j values are determined for each plastic type, j = 1,2,3. The micro proportion is calculated based on the number of decisions per object, while the macro proportion is calculated based on the average decision per class. The performance measurements refer to Table 2 for the first plastic type. The performance measure for other plastic types is determined similarly [13][15].

Table 2: Confusion Matrix for plastic-type, j = 1

		Actual		
	j	**1**	**2**	**3**
Prediction	**1**	True-Positive (TP)	False-Negative (FN)	False-Negative (FN)
	2	False-Positive (FP)	True-Negative (TN)	True-Negative (TN)
	3	False-Positive (FP)	True-Negative (TN)	True-Negative (TN)

$$Accuracy = \frac{\sum_{j=1}^{3} \frac{TP_j + TN_j}{TP_j + FP_j + FN_j + TN}}{3} \quad (3)$$

$$Recall_\mu = \frac{\sum_{j=1}^{3} TP_j}{\sum_{j=1}^{3} (TP_j + FN_j)} \quad (4)$$

$$Recall_M = \frac{\sum_{j=1}^{3} \frac{TP_j}{TP_j + FN_j}}{3} \quad (5)$$

$$Specificity_\mu = \frac{\sum_{j=1}^{3} TN_j}{\sum_{j=1}^{3} (FP_j + TN_j)} \quad (6)$$

$$Specificity_M = \frac{\sum_{j=1}^{3} \frac{TN_j}{FP_j + TN_j}}{3} \quad (7)$$

3. RESULTS AND DISCUSSION

The training and test data composition is shown in Table 3, where the data were randomly divided into five folds of comparable size [13][15]. Thus, the test data for each computation is one-fold, whereas the training data is four-fold.

Table 3: Composition of training and test data

Data	**Resampling**									
Test	**Fold 1**	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Fold 6**	**Fold 7**	**Fold 8**	**Fold 9**	**Fold 10**
PET	17	15	12	23	17	17	12	14	13	10
HDPE	13	12	14	10	15	16	17	17	17	19
PP	15	18	19	12	13	12	16	14	15	16
Sum	45	45	45	45	45	45	45	45	45	45
Train	**Except Fold 1**	**Except Fold 2**	**Except Fold 3**	**Except Fold 4**	**Except Fold 5**	**Except Fold 6**	**Except Fold 7**	**Except Fold 8**	**Except Fold 9**	**Except Fold 10**
PET	133	135	138	127	133	133	138	136	137	140
HDPE	137	138	136	140	135	134	133	133	133	131
PP	135	132	131	138	137	138	134	136	135	134
Sum	405	405	405	405	405	405	405	405	405	405

All folds except fold-1 are used as training data in the first learning model computation. Fig. 2 and Fig. 3 present the decision process in a tree structure using discretization into two and three categories, respectively.

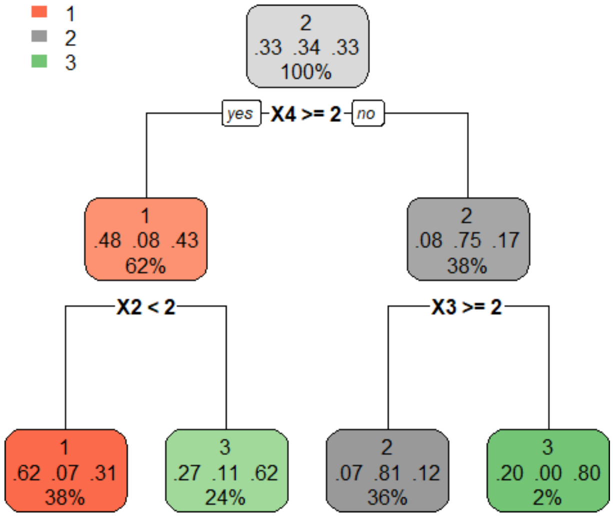

Fig. 2. Tree structure using discretization into two categories

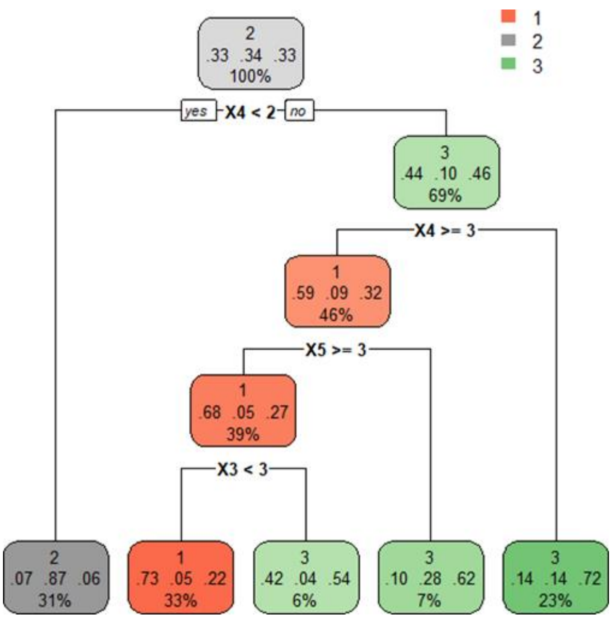

Fig. 3. Tree structure using discretization into three categories

Table 4: Performance of Plastic Waste Classification Based on Two Categories

Testing Data		**Accuracy**	**Recall$_{\mu}$**	**Recall$_{M}$**	**Specificity$_{\mu}$**	**Specificity$_{M}$**
Fold	1	74.81	62.22	62.77	81.11	80.84
	2	79.26	68.89	70.56	84.44	84.25
	3	79.26	68.89	70.09	84.44	84.37
	4	76.30	64.44	66.18	82.22	81.54
	5	74.81	62.22	62.26	81.11	80.89
	6	73.33	60.00	57.92	80.00	80.04
	7	85.19	77.78	77.21	88.89	88.72
	8	80.74	71.11	69.89	85.56	85.56
	9	86.67	80.00	79.84	90.00	90.19
	10	79.26	68.89	72.81	84.44	86.08
Average			78.96	68.44	68.95	84.22
Standard Deviation			3.15	4.72	7.10	2.35

Table 5: Performance of Plastic Waste Classification Based on Three Categories

Testing Data		Accuracy	Recall μ	Recall M	Specificity μ	Specificity M
Fold	1	85.19	77.78	78.06	88.89	88.73
	2	83.7	75.56	75.93	87.78	87.73
	3	83.7	75.56	75.75	87.78	87.74
	4	86.67	80	78.43	90	89.57
	5	77.78	66.67	64.72	83.33	82.83
	6	80.74	71.11	69.36	85.56	85.51
	7	77.78	66.67	64.83	83.33	83.25
	8	85.19	77.78	77.45	88.89	89.13
	9	83.7	75.56	75.31	87.78	87.98
	10	74.81	62.22	63.14	81.11	82.63
Average		81.93	72.89	72.30	86.45	86.51
Standard Deviation		7.34	11.00	10.55	5.50	4.31

Discretization of the predictor variables into three categories informs that the proposed decision tree model has higher performance compared to the two categories with an accuracy of 81.93 %, a recall-micro of 72.89 %, a recall-macro of 72.30 %, a specificity-micro of 86.45%, and the specificity-macro of 86.51%, respectively as presented in Table 4 and 5. The micro is determined by the number of decisions made for each object. In comparison, the macro is calculated based on the average decision made by each class. Unfortunately, this work's result is not better than Khona'ah et al. (2019), who implemented the ANNB algorithm to predict the plastic types with an accuracy of 86.67% [7]. Although the difference in prediction accuracy does not reach 1%, this work has proposed different validation techniques and more performance measures than Khona'ah et al. (2019) to show that the prediction results have low variance [7]. Therefore, better prediction performance for plastic types than our proposed method can be obtained by implementing classification methods that do not require the assumption of a multivariate Gaussian distribution and homogeneity of the covariance matrix. These methods include k-NN, decision tree, or Support Vector Machine.

4. CONCLUSION

Plastic recycling is a more environmentally friendly method of managing and reducing plastic waste that can significantly reduce land degradation, pollution, and greenhouse gas emissions. This stage is crucial because inaccurate sorting of plastic types can cause cross-contamination and increase industrial operating costs. This paper evaluates the performance of the Decision Tree model to predict the plastic-type using digital images. This model successfully predicts the plastic type. Performance measures the accuracy of 87.11 %, and the micro and macro proportion of plastic-type with correctly predicted (recall) was 91.67 % and 80.97 %, respectively. In contrast, the micro and macro proportion of the plastic type into other types predicted correctly (specificity) was 90.33 % and 90.38 %, respectively. However, superior prediction performance for plastic types can be obtained using classification methods that do not require the assumption of a multivariate Gaussian distribution and homogeneity of the covariance matrix, for example, k-NN, decision tree, or Support Vector Machine.

ACKNOWLEDGEMENT

The research/publication of this article was funded by DIPA of Public Service Agency of Universitas Sriwijaya 2021. SP DIPA-023.17.2.677515 /2021, On November 23, 2020. In accordance with the Rector's Decree Number: 0010/ UN9/ SK.LP2M.PT/2021, On April 28, 2021.

REFERENCES

[1] T. D. Moshood, G. Nawanir, F. Mahmud, F. Mohamad, M. H. Ahmad, and A. A. Ghani, "Expanding policy for biodegradable plastic products and market dynamics of bio-based plastics: Challenges and opportunities," *Sustain.*, 2021, doi: 10.3390/su13116170.

[2] R. Mülhaupt, "Green polymer chemistry and bio-based plastics: Dreams and reality," *Macromol. Chem. Phys.*, 2013, doi: 10.1002/macp.201200439.

[3] I. Muhammad Shamsuddin, "Bioplastics as Better Alternative to Petroplastics and Their Role in National Sustainability: A Review," *Adv. Biosci. Bioeng.*, 2017, doi: 10.11648/j.abb.20170504.13.

[4] R. Verma, K. S. Vinoda, M. Papireddy, and A. N. S. Gowda, "Toxic Pollutants from Plastic Waste- A Review," *Procedia Environ. Sci.*, 2016, doi: 10.1016/j.proenv.2016.07.069.

[5] J. Bobulski and M. Kubanek, "Deep Learning for Plastic Waste Classification System," *Appl. Comput. Intell. Soft Comput.*, 2021, doi: 10.1155/2021/6626948.

[6] N. Kaile, M. Lindivat, J. Elio, G. Thuestad, Q. G. Crowley, and I. A. Hoell, "Preliminary Results From Detection of Microplastics in Liquid Samples Using Flow Cytometry," *Front. Mar. Sci.*, 2020, doi: 10.3389/fmars.2020.552688.

[7] B. Khonaáh, D. Rosiliani, and I. Yani, "Identification and Clasification of Plastic Color Images based on The RGB Method," *J. Multidiscip. Eng. Sci. Technol.*, 2019.

[8] I. Yani, D. Rosiliani, B. Khona'ah, and F. A. Almahdini, "Identification and plastic type and classification of PET, HDPE, and PP using RGB method," 2020, doi: 10.1088/1757-899X/857/1/012015.

[9] I. Yani, Y. Resti, F. Burlian, and Ansyori, "Prediction of Plastic-Type for Sorting System using Fisher Discriminant Analysis," *Sci. Technol. Indones.*, 2021, doi: 10.26554/sti.2021.6.4.313-318.

[10] J. J. Rodríguez, G. Quintana, A. Bustillo, and J. Ciurana, "A decision-making tool based on decision trees for roughness prediction in face milling," *Int. J. Comput. Integr. Manuf.*, 2017, doi: 10.1080/0951192X.2016.1247991.

[11] H. Wang, J. A. Sanchez-Molina, M. Li, and F. R. Díaz, "Improving the performance of vegetable leaf wetness duration models in greenhouses using decision tree learning," *Water (Switzerland)*, 2019, doi: 10.3390/w11010158.

[12] H. nan Guo, S. biao Wu, Y. jie Tian, J. Zhang, and H. tao Liu, "Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: A review," *Bioresource Technology*. 2021, doi: 10.1016/j.biortech.2020.124114.

[13] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, 2009, doi: 10.1016/j.ipm.2009.03.002.

[14] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, 2021, doi: 10.38094/jastt20165.

[15] S. Dinesh and T. Dash, "Reliable Evaluation of Neural Network for Multiclass Classification of Real-world Data," no. 1, 2016, [Online]. Available: http://arxiv.org/abs/1612.00671.