02_Conf_2019_Study in Development of Cans Waste Classification System Based on Statistical Approaches

By Yulia Resti

Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Study in Development of Cans Waste Classification System Based on Statistical Approaches

To cite this article: Y Resti *et al* 2019 *J. Phys.: Conf. Ser.* **1198** 092004

IOP Conf. Series: Journal of Physics: Conf. Series **1198** (2019) 092004 doi:10.1088/1742-6596/1198/9/092004

Study in Development of Cans Waste Classification System Based on Statistical Approaches

Y Resti[1*], A.S. Mohruni[2], T. Rodiana[1], D.A. Zayanti[1]

[1]Mathematics, Faculty of Natural Science and Mathematics, Universitas Sriwijaya, Jl. Raya Palembang-Prabumulih Km.32 Inderalaya 30662, Ogan Ilir, Sumatera Selatan, Indonesia
[2]Mechanical Engineering, Faculty of Engineering, Universitas Sriwijaya, Jl. Raya Palembang-Prabumulih Km.32 Inderalaya 30662, Ogan Ilir, Sumatera Selatan, Indonesia

*E-mail: yulia_resti@mipa.unsri.ac.id

Abstract. A classification system is an initial stage in recycling cans technology. This study aims to build a cans waste classification system based on statistical approach using Red, Green, and Blue image. The image data size is reduced using principal component analysis, while the classification method used is K-Nearest Neighbors and multinomial regression. The results showed that the K-Nearest Neighbors method has the success rates of cans waste classification higher than Multinomial Regression.

Keyword: Cans, Classification, Statistical

1. Introduction
Recycling is a waste management technology that can reduce the occurrence of the pollution. The initial stage in recycling cans technology is sorting cans that will be recycled by type. In the process of sorting cans is automatically required a classification system. This paper aims to build a classification system of cans waste using Red, Green, and Blue (RGB) image based on statistical approaches.

The use of RGB image for classification has done many researchers with various methods. Widhiasih et al (2013) classifies star fruit based on RGB image using K-Nearest Neighbour (KNN) method in which the image is captured on the side position. The accuracy level is 80% if based on RG variables and accuracy of 91% if based on RGB variable. In another study, Nursalim et al (2014) classified graduate work using the K-NN method and compared its degree of accuracy with other methods of Naive Bayes (NB), Decision Tree (DT), Neural Network (NN) and Support Vector Machine SVM). The results showed that K-NN method has an accuracy of 83.33%. Specifically, the system is built on the variation of the angle of lighting, the type of lighting (lamp), and the speed of the conveyor belt. Principal Component Analysis (PCA) is used to reduce correlated variables in the data, while K-Nearest Neighbors (K-NN) and multinomial regression are used to classify cans by type.

2. Methodology
2.1. Principal Component Analysis (PCA)
Let X as the vector of p-variables of cans waste image poses having the covariance matrix S, A is the eigen vector of X, the linear combination of X and A,

$$Y_j = A_j^T X$$

IOP Conf. Series: Journal of Physics: Conf. Series **1198** (2019) 092004 doi:10.1088/1742-6596/1198/9/092004

is the j-th principal component if only if $var(Y_j) = A_j{}^T S A_j$ is maximum, with constraints $A_j{}^T A_j = 1$. Number of principal components in the PCA can be obtained using three ways, (1) the cumulative value of the total variance proportion, (2) correlation matrix, and (3) a scree plot.

2.2. K- Nearest Neighbors (KNN)

The k-Nearest Neighbor (KNN) method is a classification method based on the majority of the k-nearest neighbor categories. In the process of classifying the data is divided into training data and test data. The distance of each test data to each training data is calculated and retrieved as far as k data which has the closest distance then taken the majority of the group of k data which will be the group of calculated test data. In this paper, the calculation of the distance between the train data and the test data using the Euclidean Distance and Mahalanobis Distance Definition.

Suppose x_{li} is the train data for the i-th object, x_{ui} is the test data for the i-th object, n is the number of variables the image data has, S is the covariance matrix, the Euclidean Distance (d_{ij}^E) and the Mahalanobis Distance (d_{ij}^M) consecutively defined as,

$$d_{ij}^E = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2} \quad (1)$$

$$d_{ij}^M = |(x_{1i} - x_{2i})^T S^{-1} (x_{1i} - x_{2i})| \quad (2)$$

2.3. Multinomial Regression (MR)

Let N_{ij} be the random variable for i-th can waste with probability into j-th type, $\pi_j(x_i)$. The multinomial regression model was formed with stated $P(N_{i1}=1|x_i)=\pi_1(x_i)$, $P(N_{i2}=1|x_i)=\pi_2(x_i)$ and $P(N_{i3}=1|x_i)=\pi_3(x_i)$ as the following logistic functions, where cans waste type 3 as reference category (Resti et al, 2017),

$$\pi_1(x_i) = \frac{g_1}{1+g_1+g_2} \quad (3)$$

$$\pi_2(x_i) = \frac{g_2}{1+g_1+g_2} \quad (4)$$

and

$$\pi_3(x_i) = \frac{1}{1+g_1+g_2} \quad (5)$$

The $\boldsymbol{g_j}$ is defined as

$$g_j = \sum_{k=0}^{p} \exp(\beta_k x_{ijk}) \quad (6)$$

The regression parameters, β_k, can be estimated from the probability mass function using maximum likelihood estimation.

3. Result and Discussion

RGB image data of cans waste is obtained by capturing the can image placed on a conveyor belt on three poses; top, bottom, and side by using the camera connected to the computer with the help of MATLAB software. Let i be the angle of lighting, i=1,2,3,4; j be the lamp, j = 1,2, and k be the speed of conveyor belt. Analysis of Variance (Anova) of the red colour pixel for the i-th angle, the j-th lamp, and the k-th speed is presented in Table 1.

From Table 1, it is known that there is an interaction effect between the i-th angle and j-th lamp, where with the Least Significant Difference (LSD) test it is found that the most significant treatment to the red pixel value is the 2-nd angle and the 1-st lamp, where the pixel value at the 2-nd speed has a smaller variance.

IOP Conf. Series: Journal of Physics: Conf. Series **1198** (2019) 092004 doi:10.1088/1742-6596/1198/9/092004

Table 1. Anova of Red Colour Pixel

Source of variance	df	Sum of squares	Total squares	F-stat	F table 5%
treatment	7	228,2460	32,61	5,7284	
angle of lighting	3	108,6121	36,20	6,3604	4,35
lamp	1	43,6283	43,63	7,6647	5,59
angle of lighting*lamp	3	76,0056	25,34	4,4509	4,35
Kelompok	1	4,6670	4,67	0,8199	
Galat	7	39,8448	5,69		
Total (terkoreksi)	15	268,0909	17,87		

Table 2. Anova of Green Colour Pixel

Source of variance	df	Sum of squares	Total squares	F-stat	F table 5%
treatment	7	196,788	28,11	4,9656	
angle of lighting	3	114,3703	38,12	6,7338	4,35
lamp	1	14,9135	14,91	2,6342	5,59
angle of lighting*lamp	3	67,5043	22,50	3,9745	4,35
Kelompok	1	7,9617	7,96	1,4063	
Galat	7	39,6306	5,66		
Total (terkoreksi)	15	236,4186	15,76		

Table 3. Anova of Blue Colour Pixel

Source of variance	df	Sum of squares	Total squares	F-stat	F table 5%
treatment	7	154,0464	22,01	5,7556	
angle of lighting	3	95,0234	31,67	8,2841	4,35
lamp	1	3,8027	3,80	0,9945	5,59
angle of lighting*lamp	3	55,2203	18,41	4,8141	4,35
Kelompok	1	5,5920	5,59	1,4625	
Galat	7	26,7648	3,82		
Total (terkoreksi)	15	180,8112	12,05		

Table 2 shows that only the angle affects the color of green RGB image pixels. Based on the LSD test it is known that the 4th angle most significantly affects the green pixel value, where the pixel value at the 2-nd speed and the 1-st lamp has the smallest variance.

From Table 3, it is found that there is an interaction effect between the i-th angle and the j-lamp. Based on LSD test it is known that the most significant treatment to the blue pixel value is the 2-nd angle and the 1-st lamp. Then, the data of the cans image at the both 2-nd and 4-th angles for each of the 1-st lamps and the 2-nd speed are taken as further analysis data. Each of these data is then divided into training data and test data.

Suppose that the independent variable of the RGB cans waste image pixel poses consists of a red top (C_1), a red bottom (C_2), a red side (C_3), a green top (C_4), a green bottom (C_5), a green side (C_6), a blue top (C_7), a blue bottom (C_8), and a blue side (C_9). Considering the scree plot shown in Fig. 1 and the total variance proportion shown in Table 4, it was decided that the number of PCs is 2. The first PC usage was able to explain the variance of the 86.8% data. By using Minitab software obtained eigenvalues of independent variables are $\lambda_1 = 771{,}67$, $\lambda_2 = 84{,}83$, $\lambda_3 = 11{,}25$, $\lambda_4 = 6{,}99$, $\lambda_5 = 6{,}67$, $\lambda_6 = 3{,}48$, $\lambda_7 = 2{,}98$, $\lambda_8 = 0{,}90$ and $\lambda_9 = 0{,}49$. These values show that all pixel colors have the contribution is almost the same but the red color pixels have the biggest contribution. The use of a second PC capable of explaining the variance of 96.3% data also indicates that red pixels have the

IOP Conf. Series: Journal of Physics: Conf. Series **1198** (2019) 092004 doi:10.1088/1742-6596/1198/9/092004

greatest contribution, however, when viewed from their eigenvalues the values are against the green and blue pixel values.

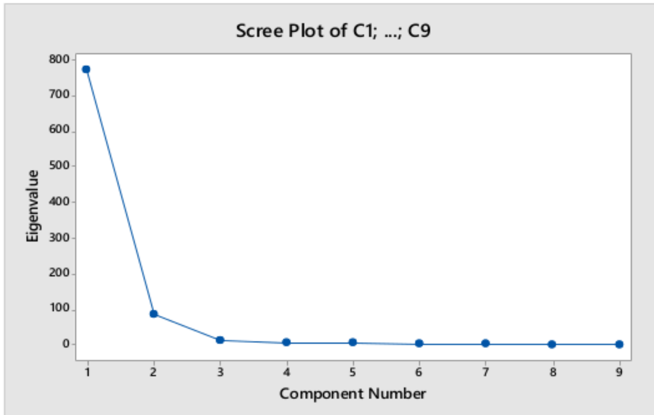

Figure 1. Scree plot between component number, p, and eigenvalue, λ_p.

Table 4. Total variance proportion

Total variance proportion 1	86,8%
Total variance proportion 1	96,3%
Total variance proportion 1	97,6%
⋮	⋮
Total variance proportion 9	100%

Table 5 present the classification result using both KNN and MR method. Each of data is divided into training data (50%) and test data (50%). For data without treatment and data with treatment of 4-nd angle, 1-st lamp, 2-rd speed, the KNN method has a higher success rate than MR method, but for data with treatment of 2-nd angle, 1-st lamp, 2-rd speed, the KNN method has a lower success rate than MR method.

Table 5. The success rate of cans waste classification using KNN and MR method

No.	Data	Classification method	Success Rate of Classification
1	without treatment	KNN	70.4%
		MR	56.0%
2	2-nd angle, 1-st lamp, 2-rd speed	KNN	59.2 %
		MR	64.8%
3	4-nd angle, 1-st lamp, 2-rd speed	KNN	51.2 %
		MR	49.6 %

4. Conclusion

This study builds a cans waste classification system based on statistical approach using Red, Green, and Blue image. Considering the scree plot and the total variance proportion, we have 2 principal components (PC). The first PC was able to explain the variance of the 86.8% data, where all pixel colors have the contribution is almost the same but the red color pixels have the biggest contribution. The second PC was able to explain the variance of the 96.3% data where all pixel colors have also the contribution is almost the same but, the values are against the green and blue pixel values. For cans waste classification system, generally, the KNN method has a higher success rate than MR method.

IOP Conf. Series: Journal of Physics: Conf. Series **1198** (2019) 092004 doi:10.1088/1742-6596/1198/9/092004

References

[1] Resti, Y., I. Yani, A.S. Mohruni, F. Burlian, and A. Amran, 2017. A Probability Approach in Cans Identification. Matec Web of Conference.A reference

[2] Whidhiasih, R.N *et al*. 2013. Klasifikasi Buah Belimbing Berdasarkan Citra Red-Green-Blue Menggunakan KNN dan LDA. Jurnal Penelitian Ilmu Komputer Universitas Pakuan Bogor: 29-35

[3] Nursalim, Suprapedi, H. Himawan. 2014. Klasifikasi Bidang Kerja Lulusan Menggunakan Algoritma *K-Nearest Neighbor*. *Jurnal Teknologi Informasi* Vol.10 No.1: 1-13.

02_Conf_2019_Study in Development of Cans Waste Classification System Based on Statistical Approaches

ORIGINALITY REPORT

21%

SIMILARITY INDEX

PRIMARY SOURCES

1	eprints.polsri.ac.id Internet	175 words — 8%
2	eprints.umpo.ac.id Internet	50 words — 2%
3	Yulia Resti, Amrifan S. Mohruni, Firmansyah Burlian, Irsyadi Yani, Ali Amran. "A probability approach in cans identification", MATEC Web of Conferences, 2017 Crossref	41 words — 2%
4	Winoto Chandra, Bambang Suprihatin, Yulia Resti. "Median-KNN Regressor-SMOTE-Tomek Links for Handling Missing and Imbalanced Data in Air Quality Prediction", Symmetry, 2023 Crossref	23 words — 1%
5	Y Harisman, A Sovia, Rahima. "The Effect of Problem Based Module for Sampling Technique Courses Toward College Student Problem Solving Behaviour; Case Study in Indonesia", Journal of Physics: Conference Series, 2021 Crossref	22 words — 1%
6	Hua Zhang. "Study on Sludge Characteristics of Lead and Zinc Smelting Acid and Sewage", IOP Conference Series: Earth and Environmental Science, 2020 Crossref	19 words — 1%

7 N Sushma Rani, P Srinivasa Rao, P Parimala. "An Efficient Statistical Computation Technique for Health Care Big Data using R", IOP Conference Series: Materials Science and Engineering, 2017 — 16 words — 1%
Crossref

8 iopscience.iop.org — 16 words — 1%
Internet

9 www.kom.tu-darmstadt.de — 15 words — 1%
Internet

10 encyclopedia.pub — 13 words — 1%
Internet

11 www.readkong.com — 13 words — 1%
Internet

12 Emre Elmaci, Fatih M. Nuroglu. "Management Relay for Assisting Distance Relays in Series Compensated Transmission Lines", Arabian Journal for Science and Engineering, 2023 — 12 words — 1%
Crossref

13 nlist.inflibnet.ac.in — 12 words — 1%
Internet

14 jesr.eng.unila.ac.id — 11 words — 1%
Internet

15 www.researchgate.net — 9 words — < 1%
Internet

16 Muhammad Akmal Fauzi, Ahmad Tri Hanuranto, Casi Setianingsih. "Intrusion Detection System using Genetic Algorithm and K-NN Algorithm on Dos Attack", — 8 words — < 1%

2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS), 2020
Crossref

17	doc.lagout.org Internet	8 words — < 1%
18	eeer.org Internet	8 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	OFF