Science and Technology Indonesia
e-ISSN:2580-4391 *p*-ISSN:2580-4405
Vol. 5, No. 4, October 2020

Research Paper

Performance of Cans Classification System for Different Conveyor Belt Speed using Naïve Bayes

Yulia Resti[1]*, Firmansyah Burlian[2], Irsyadi Yani[2]

[1] *Department of Mathematics, Faculty of Mathematics and Natural Science Universitas Sriwijaya, Sumatera Selatan, Indonesia*
[2] *Department of Machanical Engineering, Faculty of Engineering Universitas Sriwijaya, Sumatera Selatan, Indonesia*
***Corresponding author**: yulia_resti@mipa.unsri.ac.id

Abstract
The classification system in the sorting process in the can recycling industry can be made based on digital images by exploring the basic color pixel values of images such as R, G, and B as variable inputs. In real time, the classification of cans in the sorting process occurs when cans placed on a conveyor belt move at a certain speed. This paper discusses the performance of can classification systems using the Naïve Bayes method. This method can handle all types of variables, including when all variables are continuous. Two types of conveyor belts are designed to get different speeds, and all images of the cans are captured on both conveyor belts. Two models of Bayes naive are built on the basis of the different distribution assumptions; the original model (all Gaussian distributed) and the model based on the best distribution. Performance of the classification system is built by dividing data into the learning data and the testing data with a composition of 50:50 in which each data is designed into 50 groups with different percentages on each type of cans using sampling technique without replacement. The results obtained are first, the speed of the conveyor belt when capturing an image affects the pixel values of red, green, and blue and ultimately affects the results of the classification of cans. Second, not all input variables are Gaussian distributed. The classification system was built using assumption that the best distribution model for each input variable has the better average accuracy level than the model that assumes all input variables are Gaussian distributed, and the accuracy level of classification on the first speed of conveyor belt with a gear ratio of 12:30 and a diameter of 35 mm has an accuracy that is better than the other speed, both on the original model and the model based on the best distribution. However, it is necessary to test more statistical distribution models to obtain significant results.

Keywords
Classification System, Conveyor Belt Speed, Naïve Bayes

Received: 28 August 2020, Accepted: 4 October 2020
https://doi.org/10.26554/sti.2020.5.4.111-116

1. INTRODUCTION

The automation technology of an industrial system that uses intelligent computing systems has continued to develop rapidly recently (Kamboj et al. (2019); Nikhil et al. (2017); Oladapo et al. (2016); Bargal et al. (2016); Fluke (2015); Rosenblat et al. (2014)) including the automation of sorting systems in the can recycling industry that uses object classification techniques based on digital images (Resti et al. (2018); Resti et al. (2017b)). Classification of cans based on digital images of cans placed on a static conveyor belt can be seen in (Resti et al. (2019); Resti et al. (2017a); Resti (2015); Yani et al.; Yani et al. (2009)). In real time, the classification of cans in a sorting system occurs when cans placed on a conveyor belt move at a certain speed. Obtaining a higher level of accuracy becomes important in the classification system (Sin & Wang, 2019; Aronoff et al. (1982)).

Naïve Bayes is one method that is widely used in classification models (Harzevili and Alizadeh, 2018; Agarwal et al. (2015)) especially digital object classification models can be seen in (Mansour (2018); Pérez-Díaz et al. (2017); Nikhil et al. (2017); Salinas-Gutiérrez et al. (2010); Jayech and Mahjoub (2010)). This method can handle various types of input variables. When the input variables are continuous type, generally this method is built by assuming all input variables are Gaussian distributed or a combination. The concept of conditional probability as in the Bayes theorem with the naive assumptions in this method causes the calculation of posterior probabilities to be simpler (Han et al. (2011); Mitchell (1997)). For large datasets, this method often has a higher level of accuracy than other methods (Adetunji et al. (2018); Kini et al. (2015); Loan (2006)), while for small datasets this method also perform powerful classifier (Mansour, 2018).

This article discusses the performance of a can classification

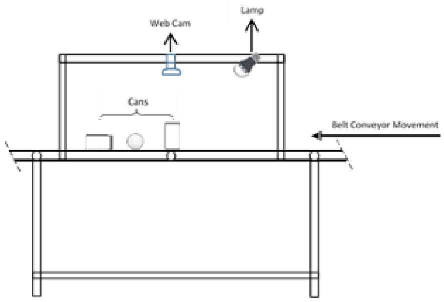

Figure 1. The cans image capturing system

system based on digital image of cans using the Naïve Bayes method. Regarding real time, the image capturing was carried out on cans placed on conveyor belts. Both types of conveyor belts are designed to get different speeds using different sizes and gear ratios, and all images of the cans are captured on both. We also propose two models of Bayes; the original model and the model based on the best distribution. In the first model, all input variables are assumed to be Gaussian distribution while in the second model, each input variable is assumed to be Gamma or Gaussian distribution according to statistical tests (Wang and Liu (2006); De Wet (1980); Stephens (1974); Chakravarty et al. (1967)). Performance of the classification system is built by dividing data into the learning data and the testing data with a composition of 50:50 in which each data is designed into 50 groups with different percentages on each type of cans using sampling technique without replacement. A highest level of accuracy is expected from the combination of these two speeds and the two models.

2. EXPERIMENTAL SECTION

2.1 Methods

The stages of this research are as follows:

1. Designing 2 types of conveyor-belts, the first using a gear with a ratio of 12:30 and diameter of 35 mm, and the second, using a gear with a ratio of 14: 30 and diameter of 42 mm. These designs produced the speed of 0.181 m/s (the first conveyor-belt) and 0.086 m/s (the second conveyor-belt) respectively.

2. Capturing images of the cans placed on the first conveyor-belt. The cans were captured using a web camera connected to a computer with the illumination of the light-emitting diode (LED) lamp set at an angle of 30° as shown in Figure 1. Then, the cans are placed on the second conveyor belt and the image capturing proses is done the same way.

Furthermore, the can image data is processed using the RGB color model with a color depth of 8 bits where the region of interest in Each image is obtained using image processing cropping techniques. Data summary of the pixel values of R, G, and B of the two data are presented in Table 1.

3. Divide the data into learning data and testing data with a composition of 50:50, where each data is designed into 50 groups with different percentages on each type of can using sampling technique without replacement. The percentage of cans in each type of the learning data and the testing data, respectively is presented in Table 2.

Table 1. Data Summary of the Pixel Values of R, G, and B

Statistics	The 1^{st} speed Input variable			The 1^{st} speed Input variable		
	R	G	B	R	G	B
Minimum	141.3	143.0	137.7	135.4	134.6	131.8
1^{st} Quartile	149.8	153.0	148.9	145.4	147.8	144.0
Median	153.3	156.2	152.2	148.1	151.1	147.5
Mean	155.4	156.4	152.7	150.4	151.5	148.3
3^{rd} Quartile	159.6	159.6	156.1	153.3	154.5	150.7
Maximum	204.8	186.2	194.8	207.6	182.5	193.4

Table 2. Design of the learning and the testing data

Percentage of Cans in Each Type							
Learning				Testing			
Group	Can Type 1^{st}	2^{nd}	3^{rd}	Group	Can Type 1^{st}	2^{nd}	3^{rd}
1	33.6	35.2	31.2	1	25.6	31.2	43.2
2	32.8	32.0	35.2	2	34.4	32.8	32.8
3	30.4	33.6	36.0	3	28.8	32.8	38.4
4	28.0	30.4	41.6	4	31.2	36.0	32.8
5	38.4	22.4	39.2	5	20.8	44.0	35.2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
50	30.4	32.0	37.6	50.0	28.8	34.4	36.8

4. Applying the Naive Bayes method [23-24] to model the can classification system. Let pixel values from each of the red, blue, and green colors for the k-th conveyor belt type be the input variables that denoted as R_k, G_k, B_k and the three types of cans are food cans, beverage cans, and non-food and beverage cans be the output variables that denoted as T_{jk}. A can is classified as a can of j-th can type, if the can has the greatest posterior probability in the j-th can type as written in (1).

$$P(T_{jk}|R_k, G_k, B_k) = \frac{P(R_k, G_k, B_k|T_{jk})P(T_{jk})}{P(R_k, G_k, B_k)} \quad (1)$$

where $P(R_k, G_k, B_k|T_{jk})$ is the likelihood function of input variables given output variable, and $P(R_k, G_k, B_k)$ is the joint probability density function.

Modeling using the Naive Bayes method consists of two assumptions; first, all input variables are assumed to be Gaussian distributions with a probability density function as in (2) for input variable R_k with parameters μ_{rk} and σ_{rk}; second, each input variable is assumed to be distributed as the best distribution model of the Gaussian distribution and Gamma distribution with a probability density function as in (3) for input variable R_k with parameters θ_{rk} and β_{rk}; based on 5 goodness of fit tests;

Kolmogorov-Simirnov [28], Cramer von Mises [29], Anderson-Darling [30], Akaike Information Criteria and Bayesian Information Criteria [31]. The first assumption is called original model (OM), while the second assumption is called the best model (BM).

$$P(R_k; \mu_{rk}, \sigma_{rk}) = \frac{1}{\sigma_{rk}\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{r_k - \mu_{rk}}{\sigma_{rk}}\right)^2\right] \quad (2)$$

$$P(R_k; \mu_{rk}, \sigma_{rk}) = \frac{1}{\beta_{rk}^{\theta_{rk}}\Gamma(\theta_{rk})} r_k^{\theta_{rk}-1} exp\left[-\left(\frac{r_k}{\beta_{rk}}\right)\right] \quad (3)$$

5. Measuring the performance of can classification for each conveyor-belt type data (first speed and second speed) and both model assumptions; original model (OM) and model based on best distribution (BM). OM assumes all input variables are Gaussian distributed while BM is a model based on the best distribution of input variables. The accuracy performance is calculated as the mean of accuracy level.

3. RESULTS AND DISCUSSION

3.1 The best distribution model of input variables

All variables from the two data are tested with 5 goodness of fit tests to determine the suitability of each variable with the Gaussian and Gamma distribution models. The results of 5 goodness of fit tests for the 1^{st} speed data are given in Table 3, while the parameters of the best distribution models are given in Table ??

The distribution model that has smaller goodness of fit value is a better model. Each of input variable has 2 - 5 tests that support it as the best model. The best distribution model of the input variables R, G, and B are all Gamma distributions, on the 2^{nd} can type are all Gaussian distributions, while on the 3^{rd} cans type are Gamma, Gaussian, and Gamma distributions, respectively.

The results of the goodness of fit tests of all input variables for each can type of the second conveyor belts speed and the parameters of the best distribution models are given in Table 5 and Table 6 successively.

Table 5 informs that on the 1^{st} can type, the best distribution model of the input variables R, G, and B are all Gamma distributions, on the 2^{nd} can type are Gamma, Gaussian, Gamma distributions, while on the 3^{rd} cans type are all Gamma distributions. In the 2^{nd} speed, at least each input variable has three tests that support it as the best model, and on average it has four tests that support it.

3.2 Performance of Classification

Table 7 shows the accuracy level of each conveyor-belt type both in the original model (OM) and the best distribution model (BM).

Each group has a different accuracy level for each conveyor-belt type both in the original model (OM) and the best distribution model (BM). To that end, classification performance is measured as the mean of the accuracy levels of the 50 groups. The variances, bias and confidence interval of the mean of the 50 groups are also presented in Table 7.

The mean of the classification accuracy level of 50 groups noted that BM has better accuracy than OM, both on the 1^{st} data speed (the 1^{st} conveyor-belt type) with a difference of 0.4%, and the 2^{nd} data speed (the 2^{nd} conveyor-belt type) with a difference of 0.1%. The variance, bias, and confidence interval of the mean in both data also show that BM has better performance than OM. This small difference in the four statistics can be caused by the variable distribution model adjusted for each input variable only two, namely Gaussian and Gamma.

Fitting the distribution of input variables to more distribution models allows a more appropriate distribution model to be obtained so that the level of accuracy can be higher. Comparison of the measurement of accuracy of the 1^{st} and 2^{nd} speeds for both OM and BM has a difference of around 6-7%, a bias difference of around 5-8%, and a confidence interval of more than 7%. These measurements show that the performance of can classification at the 1^{st} speed is better than the 2^{nd} speed at both OM and BM.

4. CONCLUSIONS

This paper proposed the performance of a can classification system based on the digital image built using 2 types of conveyor belts and 2 types of models in the Naive Bayes method to obtain the highest level of accuracy. The performance of the classification accuracy is built by dividing data into the learning data and testing data with a composition of 50:50 in which each data is designed into 50 groups with different percentages on each type of cans using resampling techniques with replacement. The results show that the classification system was built using assumption the best distribution model for each input variable has a better performance of accuracy than the model that assumes all input variables are Gaussian distributed, and the performance of accuracy on the first speed is better than the second speed, both on the original model (OM) and the model based on the best (BM) distribution. Overall, the best classification performance is owned by the Naive Bayes method which assumes the best distribution model for each input variable where image data is obtained from the capturing system with a conveyor belt speed of 0.181 m/s. Important notes from the results of this study are first, the conveyor belt speed when capturing images affects the pixel value of red, green, and blue and ultimately affects the results of the classification of cans. Second, not all input variables are Gaussian distributed. Implementation of the best statistical distribution model on the Naïve Bayes method can influence the results of classification but it is necessary to test more statistical distribution models to obtain significant results.

5. ACKNOWLEDGEMENT

This research was supported by DIPA, University of Sriwijaya, No. SP DIPA-042.01.2.400953/2019, for the Competitive Research, No. 0015 /UN9/SK.LP2M.PT/2019.

Table 3. Goodness-of-fit test for the 1^{st} speed

Input Variable	Goodness of fit	The 1^{st} cans type Gaussian	The 1^{st} cans type Gamma	The 2^{nd} cans type Gaussian	The 2^{nd} cans type Gamma	The 3^{rd} cans type Gaussian	The 3^{rd} cans type Gamma
R_1	KS	0.12	0.12	0.07	0.06	0.06	0.06
	CVM	0.20	0.17	0.06	0.07	0.04	0.04
	AD	1.47	1.28	0.55	0.58	0.27	0.26
	AIC	599.42	595.44	404.34	405.02	619.17	618.61
	BIC	604.03	600.05	409.18	409.86	624.24	623.67
G_1	KS	0.11	0.11	0.09	0.09	0.06	0.07
	CVM	0.13	0.12	0.16	0.16	0.05	0.06
	AD	1.02	0.95	1.33	1.40	0.34	0.39
	AIC	555.60	553.72	399.01	400.00	548.47	548.81
	BIC	560.21	558.33	403.84	404.84	553.54	553.87
B_1	KS	0.12	0.11	0.12	0.13	0.06	0.06
	CVM	0.32	0.27	0.16	0.16	0.06	0.06
	AD	2.20	1.87	1.06	1.13	0.40	0.40
	AIC	573.56	568.51	426.78	428.04	576.68	575.79
	BIC	578.17	573.12	431.62	432.88	581.74	580.85

Table 4. Parameter of the best distribution model for the 1^{st} speed

Input Variable	The 1^{st} cans type Parameter		The 2^{nd} cans type Parameter		The 3^{rd} cans type Parameter	
R_1	$\theta_{r_{11}}$	144.89	$\mu_{r_{21}}$	150.87	$\beta_{r_{31}}$	565.11
	$\beta_{r_{11}}$	0.91	σ_{r21}	12.49	$\beta_{r_{31}}$	3.61
G_1	$\theta_{g_{11}}$	249.84	$\mu_{g_{21}}$	154.02	$\mu_{g_{31}}$	158.03
	$\beta_{g_{11}}$	1.59	$\sigma_{g_{21}}$	2.63	$\sigma_{g_{31}}$	4.54
B_1	$\theta_{b_{11}}$	193.40	$\mu_{b_{21}}$	150.84	$\theta_{b_{31}}$	866.93
	$\beta_{b_{11}}$	1.27	$\sigma_{b_{21}}$	3.11	$\beta_{b_{31}}$	5.62

Table 5. Goodness-of-fit test for the 2^{nd} speed

Input Variable	Goodness of fit	The 1^{st} cans type Gaussian	The 1^{st} cans type Gamma	The 2^{nd} cans type Gaussian	The 2^{nd} cans type Gamma	The 3^{rd} cans type Gaussian	The 3^{rd} cans type Gamma
R_2	KS	0.13	0.12	0.07	0.07	0.11	0.11
	CVM	0.32	0.26	0.09	0.09	0.25	0.22
	AD	1.79	1.45	0.62	0.60	1.28	1.13
	AIC	603.33	598.39	460.68	460.54	627.62	625.43
	BIC	607.94	603.00	465.52	465.38	632.68	630.49
G_2	KS	0.08	0.08	0.08	0.08	0.06	0.05
	CVM	0.11	0.08	0.09	0.10	0.05	0.04
	AD	0.70	0.54	0.61	0.66	0.31	0.27
	AIC	552.90	550.84	441.79	442.09	597.78	596.59
	BIC	557.51	555.45	446.62	446.93	602.85	601.65
B_2	KS	0.15	0.14	0.07	0.07	0.08	0.08
	CVM	0.41	0.34	0.04	0.04	0.17	0.15
	AD	2.41	1.96	0.31	0.32	1.31	1.09
	AIC	576.11	570.59	445.38	445.40	608.33	604.17
	BIC	580.71	575.20	450.22	450.22	613.39	609.23

Table 6. Parameter of the best distribution model for the 2nd speed

Input Variable	The 1^{st} cans type Parameter		The 2^{nd} cans type Parameter		The 3^{rd} cans type Parameter	
R_2	θ_{r21}	132.13	θ_{r22}	1518.32	θ_{r23}	477.43
	β_{r21}	0.85	β_{r22}	10.29	β_{r23}	3.19
G_2	θ_{g21}	246.83	μ_{g22}	150.63	θ_{g23}	663.66
	β_{g21}	1.61	σ_{g22}	3.38	β_{g23}	4.40
B_2	θ_{b21}	180.19	μ_{b22}	147.95	θ_{b23}	585.61
	β_{b21}	1.20	σ_{b22}	3.46	β_{b23}	3.97

Table 7. Performance of Naive Bayes

Group	Accuracy level of classification (%)			
	1^{st} speed		2^{nd} speed	
	OM	BM	OM	BM
1	73.6	72.8	64.8	66.4
2	75.2	76.0	72.0	71.2
3	78.4	77.6	69.6	69.6
4	71.2	72.8	67.2	69.6
5	80.0	78.4	79.2	76.8
⋮	⋮	⋮	⋮	⋮
50	81.6	81.6	68.8	69.6
Mean	76.6	77.0	69.2	69.3
Variance	11.5	9.7	17.5	16.9
Biased of mean	12.2	9.5	17.2	17.0
confidence interval of mean	76.1 - 77.2	76.5 – 77.6	68.6 – 69.8	68.7 – 69.9

REFERENCES

Adetunji, A., J. Oguntoye, O. Fenwa, and N. Akande (2018). Web Document Classification Using Naïve Bayes. *Journal of Advances in Mathematics and Computer Science*; 1–11

Agarwal, S., N. Jain, and S. Dholay (2015). Adaptive testing and performance analysis using naive bayes classifier. *Procedia Computer Science*, **45**; 70–75

Aronoff, S. et al. (1982). Classification accuracy: a user approach. *Photogrammetric Engineering and Remote Sensing*, **48**(8); 1299–1307

Bargal, N., A. Deshpande, R. Kulkarni, and R. Moghe (2016). PLC based object sorting automation. *International Research Journal of Engineering and Technology, IRJET*, **3**(07)

Chakravarty, I. M., J. Roy, and R. G. Laha (1967). Handbook of methods of applied statistics

De Wet, T. (1980). Cramér-von Mises tests for independence. *Journal of Multivariate Analysis*, **10**(1); 38–50

Fluke, J. (2015). Implementing an Automated Sorting System

Han, J., J. Pei, and M. Kamber (2011). *Data mining: concepts and techniques.* Elsevier

Harzevili, N. S. and S. H. Alizadeh (2018). Mixture of latent multinomial naive Bayes classifier. *Applied Soft Computing*, **69**; 516–527

Jayech, K. and M. A. Mahjoub (2010). New approach using Bayesian Network to improve content based image classification systems. *International Journal of Computer Science Issues (IJCSI)*, **7**(6); 53

Kamboj, D., A. Diwan, et al. (2019). Development of Automatic Sorting Conveyor Belt Using PLC. *International Journal of Mechanical Engineering and Technology*, **10**(8)

Kini, M., D. Devi, and N. Chiplunkar (2015). Text mining Approach to Classify Technical Research Document using Naïve Bayes. *International Journal of Advanced Research in Computer and Communication Engineering*, **4**(7); 386–391

Loan, P. (2006). An approach of the Naive Bayes classifier for the document classification. *General Mathematics*, **14**(4); 135–138

Mansour, A. M. (2018). Texture classification using Naïve Bayes classifier. *IJCSNS Int. J. Comput. Sci. Netw. Secur*, **18**(1); 112–121

Mitchell, T. M. (1997). Machine Learning. *New York: McGraw-Hill*

Nikhil, B., S. Pramod, G. G. W. Patil, and G. S. and (2017). Automatic sorting machine. *International Journal of Advance Research and Innovative Ideas in Education, 3(3): 2254-2262*, **3**(3); 2254–2262

Oladapo, B. I., V. Balogun, A. Adeoye, C. Ijagbemi, A. S. Oluwole, I. Daniyan, A. E. Aghor, and A. P. Simeon (2016). Model design and simulation of automatic sorting machine using proximity sensor. *Engineering science and technology, an international* *journal*, **19**(3); 1452–1456

Pérez-Díaz, Á. P., R. Salinas-Gutiérrez, A. Hernández-Quintero, and O. D. Cedeño (2017). Supervised Classification Based on Copula Functions. *Res. Comput. Sci.*, **133**; 9–18

Resti, Y. (2015). Dependence in Classification of Aluminium Waste. *Journal of Physics: Conference Series*, **622**; 012052

Resti, Y., A. Mohruni, T. Rodiana, and D. Zayanti (2019). Study in Development of Cans Waste Classification System Based on Statistical Approaches. *Journal of Physics: Conference Series*, **1198**(9); 092004

Resti, Y., A. S. Mohruni, F. Burlian, I. Yani, and A. Amran (2017a). A probability approach in cans identification. *MATEC Web of Conferences*, **101**; 03012

Resti, Y., A. S. Mohruni, F. Burlian, I. Yani, and A. Amran (2018). Design of mechanical arm for an automatic sorting system of recyclable cans. *Journal of Physics: Conference Series*, **1007**; 012066

Resti, Y., S. M. Mohruni, F. Burlian, I. Yani, and A. Amran (2017b). Automation of a cans waste sorting system using the ejector system. *Modern Applied Science*, **11**(3); 48–52

Rosenblat, A., T. Kneese, and D. Boyd (2014). Understanding intelligent systems. Data and Society Working Paper, October 8. Data and Society Research Institute.

Salinas-Gutiérrez, R., A. Hernández-Aguirre, M. J. Rivera-Meraz, and E. R. Villa-Diharce (2010). Supervised probabilistic classification based on Gaussian copulas. In *Mexican International Conference on Artificial Intelligence*. Springer, pages 104–115

Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American statistical Association*, **69**(347); 730–737

Wang, Y. and Q. Liu (2006). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock–recruitment relationships. *Fisheries Research*, **77**(2); 220–225

Yani, I., M. Hannan, H. Basri, E. Scavino, and N. E. bin Ahmad Basri (2009). Detecting Object Using Combination of Sharpening and Edge Detection Method. *European Journal of Scientific Research*, **32**(1); 121–127

Yani, I., E. Scavino, M. Hannan, D. Wahab, and H. Basri (). An Automatic Sorting System for Recycling Beverage Cans using the Eigenface Algorithm. In *Proceedings of the Third International Conference on Soft Computing Technology in Civil, Structural and Environmental Engineering*. Civil-Comp Press

