E3S Web of Conferences

Proceedings

- *Information System And Technology Advancement*

Open Access

Sentiment Analysis on Twitter Using Deep Belief Network Optimized with Particle Swarm Optimization 02001

Irma Amelia Dewi and Muhammad Aldi Rizqullah

Published online: 07 February 2024

DOI: https://doi.org/10.1051/e3sconf/202448402001

Abstract | PDF (710.8 KB) | References | NASA ADS Abstract Service

Open Access

Web Accessibility Evaluation Using WCAG 2.0 and User-Centered Design Method 02002

Kurnia Ramadhan Putra, Sofia Umaroh and Nuralia Rafina

Published online: 07 February 2024

DOI: https://doi.org/10.1051/e3sconf/202448402002

Abstract | PDF (1.057 MB) | References | NASA ADS Abstract Service

Open Access

Student's Acceptance and Actual Use of E-Learning System in a Post-COVID Era Through Technology Acceptance Model 02003

Sofia Umaroh, Mira Musrini and Fachri Achmad Maulana

Published online: 07 February 2024

DOI: https://doi.org/10.1051/e3sconf/202448402003

Abstract | PDF (639.2 KB) | References | NASA ADS Abstract Service

Open Access

Performance Analysis of Oversampling and Undersampling on Telco Churn Data Using Naive Bayes, SVM And Random Forest Methods 02004

Muhammad Daffa Nur Alif and Nur Fitrianti Fahrudin

Published online: 07 February 2024

DOI: https://doi.org/10.1051/e3sconf/202448402004

Abstract | PDF (733.0 KB) | References | NASA ADS Abstract Service

Open Access

Redesigning UI/UX Of Itenas Online Graduation Registration Website Using Lean UX And Post-Study System Usability Questionnaire 02005

Kurnia Ramadhan Putra and Diky Akmal Fauzi

Published online: 07 February 2024

DOI: https://doi.org/10.1051/e3sconf/202448402005

Abstract | PDF (1.925 MB) | References | NASA ADS Abstract Service

Open Access

An IT Asset Governance Model Design Using COBIT 2019 And ITIL V4 Framework at BKU Itenas 02006

Rayisa Yudha Pratama and Sofia Umaroh

Published online: 07 February 2024

DOI: https://doi.org/10.1051/e3sconf/202448402006

Abstract | PDF (1.178 MB) | References | NASA ADS Abstract Service

Open Access

Glaucoma Identification Using Relevance Vector Machine 02007

Yusup Miftahuddin and Alfian Ichsan

Published online: 07 February 2024

DOI: https://doi.org/10.1051/e3sconf/202448402007

Abstract | PDF (557.4 KB) | References | NASA ADS Abstract Service

Open Access

Comparison of K-Medoids and K-Means Algorithms in Segmenting Customers based on RFM Criteria 02008

Nur Fitrianti Fahrudin and Rini Rindiyani

Published online: 07 February 2024

DOI: https://doi.org/10.1051/e3sconf/202448402008

Abstract | PDF (1.186 MB) | References | NASA ADS Abstract Service

Open Access

Classification of Roasted Coffee Beans with Principal Component Analysis and Random Forest 02009

Yusup Miftahuddin and Refda Rais

Published online: 07 February 2024

DOI: https://doi.org/10.1051/e3sconf/202448402009

Abstract | PDF (724.8 KB) | References | NASA ADS Abstract Service

Open Access

Implementation and Optimization Of Inception Resnet-v2 with Data Balancing (Case Study Of Lung Disease Classification) 02010

Galih Ashari Rakhmat and Aditya Muhammad Yudantira

Published online: 07 February 2024

DOI: https://doi.org/10.1051/e3sconf/202448402010

Abstract | PDF (631.5 KB) | References | NASA ADS Abstract Service

Open Access

Bandwidth Limitation Based on Content Classification Using Queue Trees and Hierarchical Token Buckets 02011

Lisa Kristiana and Alfan Zulfian Azmi

Published online: 07 February 2024

DOI: https://doi.org/10.1051/e3sconf/202448402011

Abstract | PDF (1.651 MB) | References | NASA ADS Abstract Service

Open Access

Analyzing of Modified DES-64 Algorithm for Securing the Data on File Storage Systems 02012

Asep Rizal Nurjaman and Kurnia Ramadhan Putra

Published online: 07 February 2024

DOI: https://doi.org/10.1051/e3sconf/202448402012

Abstract | PDF (3.744 MB) | References | NASA ADS Abstract Service

Editorial board

Classification of Roasted Coffee Beans with Principal Component Analysis and Random Forest

Yusup Miftahuddin[1,*], *Refda* Rais[1]

[1]Department of Informatics, Institut Teknologi Nasional Bandung, Indonesia

Abstract. Roasted Robusta coffee beans originating from Toraja and Ciwidey, Indonesia have the same shape, color and characteristics of the coffee beans. There is often an error in recognizing coffee bean varieties just by looking with the naked eye. In terms of image classification, Random Forest has stable performance and performance. This research aims to classify coffee beans based on images of roasted coffee beans using the Random Forest and PCA algorithms. The steps taken were collecting Robusta and Toraja datasets which were taken manually and given 2000 labels. Data collection was carried out using a light intensity of 300 lux and a distance of 15 cm. The data was preprocessed, variables were reduced using PCA and Random Forest in forming the model. Testing uses 1600 training data and 400 test data to measure RF performance by changing the parameters of the number of data sets and the n value in Random Forest with the highest accuracy value of 98.05% with a value of n = 20 and 360 data tested.

1 Introduction

Indonesia occupies ranking the fourth largest coffee producing country in the world is behind Brazil, Vietnam, and Colombia [1]. The types of coffee beans grown in Indonesia are robusta and arabica coffee [2]. Arabica coffee nor robusta own diverse varieties based on origin the coffee area planted or known with single origin [3] name. In Indonesia there are a number of Arabica coffee varieties nor robusta like Madheling from Sumatra, Typica from Garut, Kopyol from Bali, Catimor from Aceh and still Lots Again varieties found in Indonesia [4]. Robusta coffee is produced coffee more Lots compared to arabica because of robusta coffee more stand to leaf [5] rust disease. Every variety coffee beans have different characteristics including covers color, shape, texture, taste, and aroma depend from origin area the coffee beans planted [6]. Identification type done coffee beans in a manner visible eye No easy for distinguished for common people so that needed skill special, even owner coffee shop or coffee farmers yet Of course capable for recognize type coffee beans with only see in a manner visible eye [7].

*Corresponding author: yusufm@itenas.ac.id

Research on coffee bean classification has been carried out by previous researchers, namely the Implementation of the Linear Discriminant Method Analysis for Classification of Coffee Beans with use extraction HSV color and extraction form use feature eccentricity, area and yield parameters accuracy by 84% [8]. Another study, namely the Introduction of Arabica Coffee Varieties Based on Shape by taking textural features on coffee beans using statistical calculations and performing feature extraction based on the shape of the coffee beans produces an accuracy of 80% [9]. Study other that is NIRS application and Principal Component Analysis for Detecting Origin of Arabica Coffee Beans with use method correct spectrum that is Pretreatment Multiplicative Correction (MSC) succeed classify coffee beans from Province of Aceh and Province of Bali with level 100% success [10].

Extraction feature use Principal Component Analysis with 95% data variation can produce accuracy by 100% compared with extraction feature use Wavelets Daubechies with accuracy of 98.611 % in research conducted by [11] Random Forest Method applied to the classification process image own 92.98% accuracy compared with SIFT method on batik [12] classification and application Random Forest with extraction characteristic use GLCM get accuracy only 65% on identification disease paddy [13]. Random Forest Method can reduce amount the variable being measured and constituting method popular [14] classification. Random Forest own accuracy better compared to with method tree decision other [15].

Study This apply method Principal Component Analysis and Random Forest on classification type coffee beans with objective for measure level accuracy method Principal Component Analysis (PCA) and Random Forest (RF) in classify type coffee beans. Coffee bean image for training data taken manually with _ distance of 15 cm from camera in jpg format and size 50 x 50 pixels. Training data 1600 images _ coffee beans with composition of 800 Ciwidey Robusta coffee beans and 800 Toraja Robusta coffee beans.

2 Material & Method

To support this research, measuring the accuracy of the Random Forest method in the classification of roasted coffee beans has the following stages.

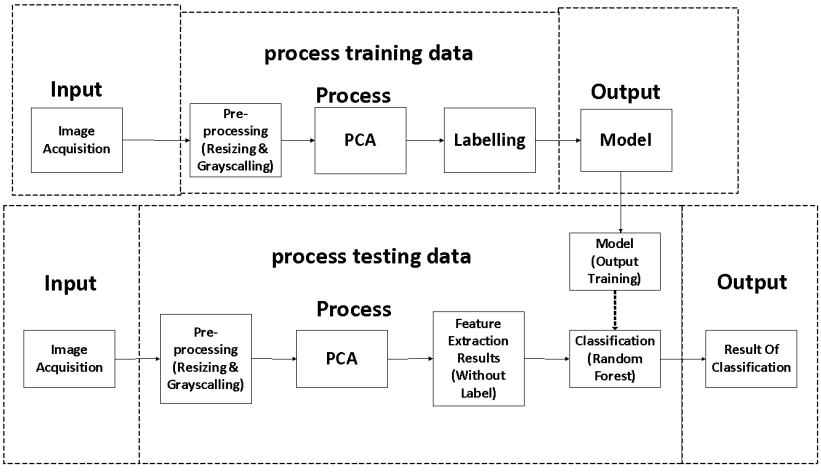

Fig 1. Research Stages

2.1 Training Data and Test Data

The dataset used is coffee bean images consisting of Ciwidey robusta and Toraja robusta with an image size of 50x50 pixels with a total of 2000 images consisting of 1600 training data and 400 test data. An example of a fundus image can be seen in Figure 2.

Fig. 2 Coffee Beans Image

2.2 Preprocessing

This study uses the stages of resizing and grayscalling. The image is resized to 50x50 pixels and the grayscaling process converts the RGB image into grayscale, as in

$$Grayscale = 0.299R + 0.587G + 0.114B \quad (1)$$

2.3 Principal Component Analysis

Principal Component Analysis is method used for simplifying a complex data, with method transform linear so that formed new coordinates with variance maximum [16]. Principal Component Analysis will produce Principal Component obtained from decomposition eigenvalue and eigenvector from covariance matrix [17]. Steps of Principal's algorithm Components Analysis is as follows:

1. Calculating *the mean* ($\overline{X}$) of the data in each dimension using equation (2)

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad (2)$$

2. Calculating *covariances matrix* (C_x) using equation (3)

$$C_x = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T \quad (3)$$

3. Calculating *the eigenvector* (v_m) and *eigenvalue* (λ_m) of *the covariance matrix* using equation (4)

$$C_x v_m = \lambda_m v_m \quad (4)$$

4. Sort *the eigenvalues* in *descending order. Principal Component* is a series *of eigenvectors* according to the order of *the eigenvalues* in stage 4.

5. Generate new datasets.

2.4 Random Forest

Random Forest is a classification method that is carried out by developing a decision method tree based on random attribute selection at each node to determine the classification [18]. Random Forest have a selection process capable features take feature best so that can increase performance to the classification [19] model. Formula calculation Random Forest can see in equation (5).

$$Gini\,(p) = 1 - p^2 - q^2 \qquad (5)$$

The gain value will be converted into a class label (0 or 1), which will determine whether the data is classified as Ciwidey robusta or Toraja robusta.

3 Results and Discussion

Random Forest testing on classification was carried out 60 times with a test scheme with a total test data of 200, 240, 280, 320, 360, and 400, and N Random Forest values of 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100. Based on these tests, the following data is obtained.

Table 1. Result Testing

Test result		**Amount of data**					
		200	**240**	**280**	**320**	**360**	**400**
Value of N RF	**10**	94.00 %	89.16 %	88.21 %	93.75 %	94.72 %	95.00 %
	20	95.50 %	95.00 %	96.78 %	97.19 %	98.05 %	97.00 %
	30	97.00 %	95.83 %	96.07 %	95.94 %	95.83 %	96.00 %
	40	97.00 %	95.83 %	96.43 %	96.25 %	96.67 %	97.00 %
	50	96.00 %	96.67 %	97.50 %	96.56 %	96.11 %	96.25 %
	60	96.50 %	96.25 %	96.78 %	96.87 %	95.55 %	96.25 %
	70	96.50 %	97.50 %	97.50 %	96.56 %	96.94 %	95.75 %
	80	98.00 %	95.83 %	97.50 %	96.25 %	96.67 %	96.00 %
	90	97.50 %	95.41 %	96.43 %	97.50 %	97.22 %	97.50 %
	100	97.00 %	96.25 %	97.14 %	97.50 %	96.94 %	97.25 %

Table 1 shows the highest average accuracy value with an accuracy of 98.05% with an N Random Forest value of 20. From testing with data of 200 to 400 images with an N Random Forest value of 10 to 100, a system performance graph can be created from the N Random Forest value to the amount of data in Figure 3.

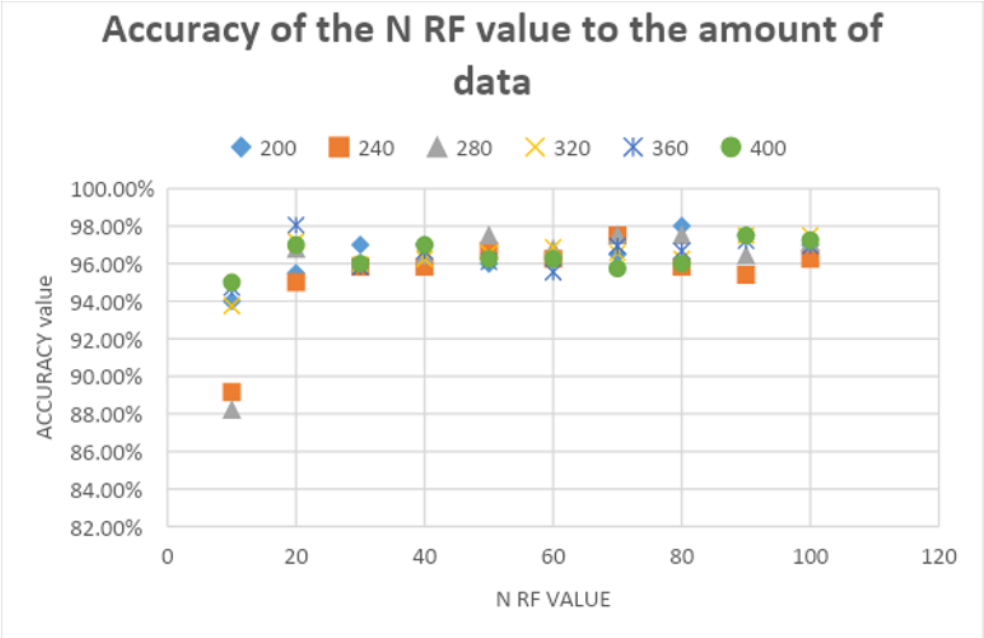

Fig. 3 System Performance

4 Conclusion

This study aims to identify and classify the types of Robusta Ciwidey or Robusta Toraja coffee beans. Starting with the preprocessing stage, namely greyscaling and resizing, feature extraction with Principal Component Analysis and classification with Random Forest. The Principal Component Analysis method is used to extract features on training data and test data using Principal Component Analysis with 95% data variation and the Random Forest classification method for classifying coffee beans. The amount of data used is 2000 image data which is divided into 1000 images of Ciwidey coffee beans and 1000 images of Toraja coffee beans with a composition of 800 training images and 200 test images. The test results with the parameters of the number of data sets and the N value in Random Forest obtained the highest accuracy results of 98.05% with a value of N = 20 and 360 data tested. Data collection was carried out using a light intensity of 300 lux and a distance of 15cm.

In addition, the researcher would like to acknowledge the funding support from LPPM Institut Teknologi Nasional Bandung.

REFERENCES

[1] K. Budiharjono, & W.M. Fahmi, Jurnal Ilmiah Mahasiswa AGROINFO GALUH, ***7*(2)**, 373-379 (2020).
[2] E.D. Martauli, Journal of Agribusiness Sciences, **1(2)**, 112-120.
[3] B.E. Cahyono, A.T. Nugroho, & I.W. Maulidina, TEKNOTAN, **16(3)**, 191-196 (2022).
[4] Rawansyah, R.A. Asmara, & T.A. Heryanto, *Klasifikasi Varietas Biji Kopi Arabika Menggunakan Ekstraksi Bentuk dan Tekstur*, in Malang: SEMINAR INFORMATIKA APLIKATIF POLINEMA (SIAP) (2019)
[5] Y. Prastyaningsih, & W. Kusrini, Jurnal Inovtek Polbeng, **6(2)**, 222-233 (2021).
[6] M.A. Nugroho, & M.M. Sebatubun, JOISM: Jurnal Of Information System Management, **1(2)**, 1-5 (2020).
[7] Y. Prastyaningsih, A. Noor, & A. Supriyanto, Computer Science and Informatics Journal, **3(2)**, 105-116 (2020).
[8] R.D. Yunita, C. Rozikin, M. Jajuli, M. Jurnal Teknologi Informatika dan Komputer MH. Thamrin, **8(1)**, 27-39 (2022).
[9] M.M. Sebatubun, & E.H. Pujiarini, Jurnal Informatika dan Komputer, **3(2)**, 60-68 (2018).
[10] S. Ramadhan, A.A. Munawar, & D. Nurba, Jurnal Ilmiah Mahasiswa Pertanian Unsyiah, **1(1)**, 954-960 (2016).
[11] R.D. Intan, & E.M. Imah, Jurnal Masyarakat Informatika, **6(11)**, 46-54 (2015).
[12] S. Devella, Yohannes, & F.N. Rahmawati, Jurnal Teknik Informatika dan Sistem Informasi, **7(2)**, 310-320 (2020).
[13] Y. Miftahuddin, S. Umaroh, & A.M. Yamani, MIND (Multimedia Artificial Intellegent Networking Database) Journal, **7(1)**, 37-50 (2022).
[14] J.L. Spesier, M.E. Miller, J. Tooze, E. Ip, Expert Systems with Applications, **134**, 93-101 (2019).
[15] M. Schonlau, & R.Y. Zou, The Stata Journal, **20(1)**, 3-29 (2020).
[16] J. Pardede, D.B. Utami, & A.C. Rochman, Jurnal Teknik Informatika dan Sistem Informasi, **3(3)**, 565-574 (2017).
[17] D. Hediyati, & I.M. Suartana, Journal Information Engineering and Educational Technology, **5(2)**, 49-54 (2021).
[18] D.A. Nugraha, & A.S. Wiguna, Journal of Computer, Information System & Technology Management, **3**, 24-30 (2020).
[19] R. Supriyadi, W. Gata, N. Maulidah, A. Fauzi, **13(2)**, 67-75 (2020).