05_JURNAL_2022_Identification of Corn Plant Diseases and Pests Based on Digital Images using Multinomial Naïve Bayes and K-Nearest Neighbor

By Yulia Resti

Science and Technology Indonesia
e-ISSN:2580-4391 p-ISSN:2580-4405
Vol. 7, No. 1, January 2022

Research Paper

Identification of Corn Plant Diseases and Pests Based on Digital Images using Multinomial Naïve Bayes and K-Nearest Neighbor

Yulia Resti[1]*, Chandra Irsan[2], Mega Tiara Putri[1], Irsyadi Yani[3], Anshori[4], Bambang Suprihatin[1]

[1]Department of Mathematics, Faculty of Mathematics and Natural Science, Sriwijaya University, Palembang, 30139, Indonesia
[2]Study Program of Plant Protection, Department of Plant Pest and Disease, Faculty of Agriculture, Sriwijaya University, Palembang, 30139, Indonesia
[3]Department of Mechanical Engineering, Faculty of Engineering, Sriwijaya University, Palembang, 30139, Indonesia
[4]Department of Electronics, Faculty of Engineering, Sriwijaya University, Palembang, 30139, Indonesia
*Corresponding author: yulia_resti@mipa.unsri.ac.id

Abstract
Statistical machine learning has developed into integral components of contemporary scientific methodology. This integration provides automated procedures for predicting phenomena, case diagnosis, or object identification based on previous observations, uncovering patterns underlying data, and providing insights into the problem. Identification of corn plant diseases and pests using it has become popular recently. Corn (*Zea mays L*) is one of the essential carbohydrate-producing foodstuffs besides wheat and rice. Corn plants are sensitive to pests and diseases, resulting in a decrease in the quantity and quality of the production. Eradicate pests and diseases according to their type is a solution to overcome the problem of disease in corn plants. This research aims to identify corn plant diseases and pests based on the digital image using the Multinomial Naïve Bayes and K-Nearest Neighbor methods. The data used consisted of 761 digital images with six classes of corn plants disease and pest. The investigation shows that the K-Nearest Neighbor method has a better predictive performance than the Multinomial Naïve Bayes (MNB) method. The MNB method with two categories has an accuracy level of 92.72%, a precision level of 79.88%, a recall level of 79.24%, F_1-score 78.17%, kappa 72.44%, and AUC 71.91%. Simultaneously, the K-Nearest Neighbor approach with k=3 has an accuracy of 99.54 %, a precision of 88.57%, recall 94.38%, F_1-score 93.59%, kappa 94.30%, and AUC 95.45%.

Keywords
Corn Plant Diseases and Pests, K-Nearest Neighbour, Naïve Bayes

Received: 10 August 2021, Accepted: 21 November 2021
https://doi.org/10.26554/sti.2022.7.1.29-35

1. INTRODUCTION

Data analysis and machine learning, known as statistical machine learning, have emerged as critical components of modern scientific practice. Their integration enables automated techniques for predicting events, diagnosing phenomena, identifying objects based on prior observations, revealing hidden patterns in the data, and providing insight into the problem. Naïve Bayes and K-Nearest Neighbor (KNN) are popular statistical learning methods to classify or identify objects (Rukmawan et al., 2021; Alsafy et al., 2014), mainly based on digital images (Umar et al., 2020; Hsu et al., 2017). The performance of both methods is satisfactory. For example, Srianto and Mulyanto (2016) classify soil suitable for planting teak trees. The obtained performance for the KNN is 96.66% accuracy, 95.45% precision, and 98.63% recall, while for the Naïve Bayes is 82.63% accuracy, 84.57% precision, and 82.02% recall. Another research is proposed by Rukmawan et al. (2021) to classy cerebral infarction. The obtained accuracy for the KNN is 91% and 97% using the Naïve Bayes method.

Corn (*Zea mays L*) is one of the foodstuffs whose productivity is threatened. For corn plants to grow well, the planting process requires adequate rainfall and an irrigation system. Nevertheless, from seed to corn ready to harvest, corn plants are sensitive to diseases and pests in their growth cycle. They can reduce the amount and quality of production. Therefore, early identification of both can reduce the risk of further damage to crops so that the quality and quantity of production can be maintained.

The implementation of statistical machine learning in identifying diseases of corn plant using digital image data has been popular recently (Ngugi et al., 2021; Xian and Ngadiran, 2021; Syarief and Setiawan, 2020; Panigrahi et al., 2020; Sibiya and Sumbwanyambe, 2019; Kusumo et al., 2019; Mengistu et al., 2018). Digital image processing using the red, green, and blue (RGB) color space model is the most informative feature in de-

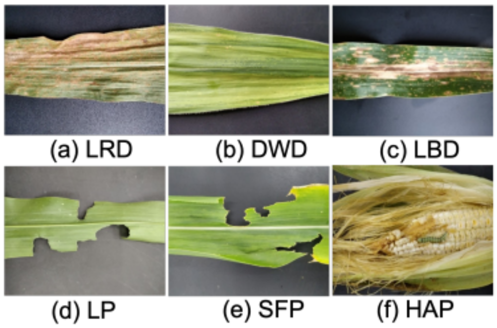

Figure 1. Digital Image of Corn Plant Disease and Pest Class

tecting corn plant diseases compared to other features such as scale-invariant feature transform (SIFT), strong feature acceleration (SURF), Oriented FAST, rotated BRIEF (ORB), and object detectors such as oriented gradient histogram (HOG). Furthermore, this property has the highest accuracy in most machine learning approaches (Kusumo et al., 2019).

However, most research that implements statistical machine learning only identifies diseases that attack corn plants but have not identified pests. This research aims to identify corn plant diseases and pests using a digital image as a database. The image is processed using an RGB color space model. The proposed statistical learning methods for identification tasks are Multinomial Naïve Bayes and KNN. The Multinomial Naïve Bayes method is a type of Naïve Bayes method that can be used as an alternative if the assumption of the Gaussian distribution of the predictor variable is not fulfilled.

2. EXPERIMENTAL SECTION

2.1 Data

The study used 761 digital images of disease and pests of corn plants to collect data. Between January and March 2021, digital images were taken with a 12 MP smartphone camera. The captures took place in corn plantations near the University of Sriwijaya, specifically in Tanjung Seteko, Tanjung Baru, and Tanjung Putus, all of which are located in South Sumatra's Ogan Ilir Regency. The data set consists of six classes; leaf rust disease (LRD), downy mildew disease (DWD), and leaf blight disease (LBD), as well as Locusta pests (LP), Spodoptera Frugiperda pest (SFP), and Heliotis Armigera pest (HAP) as presented in Figure 1(a–f).

The first three classes in Figure 1(a–c) are the types of diseases that often attack corn plants. These three diseases attack corn plants, and each class shows a different color combination. The last three classes in Figure 1(d–f) are the types of pests that often attack corn plants. The first two pests attack the corn leaves, while the last class attacks the corn fruit. In the LP image, it appears that the leaf shape is not intact, but the leaves are still green. In the SFP image, apart from incomplete leaves, it also appears that there is a yellowish color to the leaves, while in the HAP image, it appears that the corn fruit is not intact.

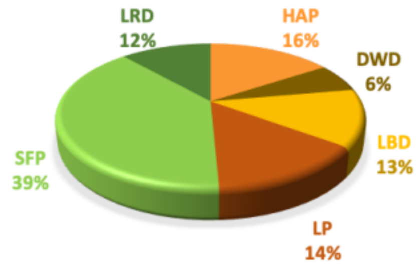

Figure 2. Class Composition of Corn Plant Disease and Pest

The data in Figure 2 shows the composition of the six classes of diseases and pests in maize, where 35% is disease data, and 65% is pest data. The most common corn plant disease was LRD as much as 16%, and corn plant pests were dominated by SFP types as much as 39%.

2.2 Method

We propose nonparametric and parametric methods from statistical machine learning to identify diseases and pests of the corn plant. Both are supervised learning. K-Nearest Neighbor (KNN) is the simplest nonparametric method of all machine learning methods. This method uses the concept of distance that similar samples are generally located close together. Thus, the value of k represents the number of neighbors or data closest to the observation. This method saves all training samples and does not build a classifier until a new sample that does not have a class needs to be classified (Han et al., 2011).

On the other hand, Naïve Bayes is a parametric method that builds a probabilistic data model following some assumptions. NB is the simple statistical Bayesian method. The name naïve is attached to the NB method because this method assumes that all predictor variables are not mutually correlated or class conditional probability is independent. This method also assumes that the predictor variables have Gaussian distribution. If the Gaussian distribution assumption is not fulfilled, the Naïve Bayes method refers to Multinomial Naïve Bayes (Chen and Fu, 2018; Kresnawati et al., 2021; Pan et al., 2018; Resti et al., 2021).

For Naïve Bayes method, suppose $X_1,...,X_w$ are predictor variables to predict the class of disease and pest of the corn plant. In the multinomial naïve Bayes method, a digital image with w-predictor variables is predicted as the j-th class of disease and pest of the corn plant (S_j) if the class of the image has a maximum posterior probability as written in Equation 1.

$$\begin{aligned}\arg\max P(S_j|X_1,\ldots,X_w) &= \arg\max \frac{P(S_j)P(X_1,\ldots,X_w|S_j)}{P(X_1,\ldots,X_w)} \\ &= \arg\max P(S_j)P(X_1,\ldots,X_w|S_j)\end{aligned} \tag{1}$$

Table 1. Confusion Matrix for the first class of disease and pest corn plant

		Prediction Class					
		LRD	DWD	LBD	LP	SFP	HAP
Actual class	LRD	TP	FN	FN	FN	FN	FN
	DWD	FP	TN	TN	TN	TN	TN
	LBD	FP	TN	TN	TN	TN	TN
	LP	FP	TN	TN	TN	TN	TN
	SFP	FP	TN	TN	TN	TN	TN
	HAP	FP	TN	TN	TN	TN	TN

Table 2. Composition of Learning and Testing Data

Data	Class					
	1 (LRD)	2 (DWD)	3 (LBD)	4 (LP)	5 (SFP)	6 (HAP)
Learning	39	78	70	86	238	96
Testing	10	20	18	22	60	24
Total	49	98	88	108	298	120

$$\arg\max P(S_j|X_1, \ldots, X_w = \arg\max P(S_j)\Pi_{d=1}^{w} P(X_d|C_j) \quad (2)$$

Equations 1 become 2 because the denominator is a constant and independent assumption of conditional probability between the variables. Each of the prior probability $P(S_j)$ and probability $P(X_d|S_j)$ in 2 is defined as,

$$P(S_j) = \frac{\sum_{d=1}^{w} n(X_d|S_j) + 1}{n + s} \quad (3)$$

$$P(X_d|S_j) = \frac{\sum_{c}^{m} n_c(X_d|S_j) + 1}{n(X_d|S_j) + m} \quad (4)$$

For each class, $n_c(X_d|S_j)$ is the number of images of the j-th class in a variable X_d with category c, $n_c(X_d|C_j)$ is the number of the image in all variables X, $n(S_j)$ is the number of images of the j-th class, m is the number of categories in the variable X_d and s is the total of corn plant disease and pest classes.

It is essential to test Gaussian assumptions before implementing the Naïve Bayes method so that interpretation and inference are reliable or valid. There are three general ways to check Gaussian assumptions; Q-Q plots, histograms, and numerical methods (statistical tests), but the last way is a more formal procedure. There are a large number of normality tests available in the literature, but the tests that are often used because they are powerful are Kolmogorov–Smirnov, Anderson-Darling, Shapiro-Wilk Razali et al. (2011), Cramer von Mises Arnastauskaitė et al. (2021), and Jarque-Bera. Kolmogorov-Smirnov, Cramer von Mises, and Anderson-Darling are tests based on the empirical distribution function (EDF). The test compares the estimated EDF based on the data with the Gaussian distribution's cumulative distribution function (CDF) to see a good match between the two. The Kolmogorov–Smirnov and the Cramer von Mises test are appropriate when the hypothesized Gaussian distribution parameters are fully known (Arnastauskaitė et al., 2021; Razali et al., 2011). Anderson-Darling test assesses whether the sample comes from a specific distribution (Adefisoye et al., 2016; Razali et al., 2011; Sibiya and Sumbwanyambe, 2019; Umar et al., 2020). In this research, the distribution in question is the Gaussian distribution, according to the assumptions of the Naïve Bayes method. Shapiro-Wilk and Jarque-Bera are tests based on skewness and kurtosis. Jarque-Bera test is based on the sample skewness and sample kurtosis, which uses the Lagrange multiplier procedure on the Pearson family of distributions to obtain tests for normality (Adefisoye et al., 2016). The Shapiro-Wilk test can detect deviations from the Gaussian distribution due to skewness, kurtosis, or both (Adefisoye et al., 2016; Arnastauskaitė et al., 2021; Razali et al., 2011).

The null hypothesis of the inference is that the predictor variable follows a Gaussian distribution. The hypothesis is rejected if the p-value is smaller than the significant level of 5%. Suppose x_i is the i-th digital image pixel value for the predictor variable X_i, $F(x_i)$ is the cumulative distribution function, $F(z_i)$ is the standard cumulative normal distribution function Z_i and n is the sample size. Kolmogorov-Smirnov (KS), Cramer von Mises (CvM), and Anderson-Darling tests statistics are shown by (Adefisoye et al., 2016; Arnastauskaitė et al., 2021; Razali et al., 2011; Jäntschi and Bolboacă, 2018).

Table 3. The Gaussian distribution assumption test

Test	R		G		B	
	stat	p-value	stat	p-value	stat	p-value
Kolmogorov-Smirnov	0.05	3.86×10^{-3}	0.12	2.20×10^{-16}	0.07	2.34×10^{-6}
Cramer von Mises	0.34	1.11×10^{-4}	2.19	7.37×10^{-10}	0.55	1.01×10^{-6}
Anderson-Darling	2.31	7.22×10^{-6}	13.16	2.20×10^{-16}	3.31	2.78×10^{-8}
Shapiro-Wilk	0.98	2.56×10^{-5}	0.93	1.81×10^{-15}	0.98	3.11×10^{-7}
Jarque-Bera	14.78	2.50×10^{-3}	109.35	2.20×10^{-16}	20.06	2.00×10^{-3}

Table 4. Discretization of Learning Data

Category	Interval		
	R	G	B
1	79.10 – 95.60	82.69 – 99.72	36.63 – 55.12
2	95.61 – 112.11	99.73 – 116.74	55.12 – 73.61
3	112.12 – 128.62	116.75 – 133.77	73.62 – 92.09
4	128.62 – 145.13	133.78 – 150.80	92.10 – 110.58
5	145.14 – 161.64	150.81 – 167.82	110.59 – 129.07

$$KS = \max_{1 \le i \le n} \left(|F(z_i) - F_{ni-1}(x_i)|, |F(z_i) - F(x_i)| \right) \quad (5)$$

$$\text{CvM} = \frac{1}{12n} + \sum_{i=1}^{n} \left(F(x_i) - \frac{2i-1}{2n} \right)^2 \quad (6)$$

$$\text{AD} = -n - \frac{1}{n} \sum_{i=0}^{n} (2i-1) \left(ln(F(x_i)) + ln(1 - F(x_{n-i+1})) \right) \quad (7)$$

Let e= $(e_1, e_2, \ldots, e_n)^T$ be the vector of the expected values of the order statistics of independent and identically distributed random variables sampled from the standard Gaussian distribution, and S be the covariance matrix of those order statistics. The constans a_i are defined as

$$(a_1, a_2, \ldots, a_n) = \frac{e^T S^{-1}}{(e^T S^{-1} S^{-1})^{\frac{1}{2}}} \quad (8)$$

The Shapiro-Wilk (SW) test is obtained using (Adefisoye et al., 2016),

$$\text{SW} = \frac{1}{D} \left(\sum_{i-1}^{n} a_i (x_{n-1+i} - x_i) \right)^2 \quad (9)$$

where

$$\text{D} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad (10)$$

Let m_2, m_3, and m_4 be the second, third, and fourth central moments. The equations b_1 and b_2 are written as,

$$b_1 = \frac{m_3^2}{m_2^3} \quad (11)$$

$$b_2 = \frac{m_4}{m_2^3} \quad (12)$$

So the Jarque-Bera (JB) test is defined as,

$$JB = n \left(\frac{b_1}{6} + \frac{(b_2 - 3)^2}{24} \right) \quad (13)$$

Furthermore, if the predictor variables are not Gaussian distribution, this work implements the discretization process as formulated as SAS Institute Inc. (1999),

$$X_d = \text{Range}(X_d) + X_d^o \quad (14)$$

where,

$$\text{Range}(X_d) = \frac{\max(X_d^o) - \min(X_d^o)}{c(X_d)} \quad (15)$$

$X_d{}^o$ be the d-th predictor variable which represents the color pixel values in the interval scale. Variable X_d is variable $X_d{}^o$ which is discretized as much as $c(X_d)$ by using 8.

In the KNN method, a digital image with w-predictor variables is predicted as the j-th class of disease and pest of corn plant if the image in the j-th class has the closest Euclidean distance to k-neighbour. Let x_{ia} be the i-th digital image pixel value X_i in training data and x_{ib} be the i-th digital image pixel value X_i in testing data, the Euclidean distance between x_{ia} and x_{ib} is defined as Han et al. (2011),

$$d_j(x_{ia}, x_{ib}) = \sqrt{\sum_{i=1}^{n}(x_{ia} - x_{ib})^2} \tag{16}$$

Next, the evaluation of methods' performance to predict the class of corn plant disease and pest use measures accuracy, precision, recall, kappa, and AUC (Dinesh and Dash, 2016; Mishra et al., 2016; Karthik and Abhishek, 2019; Sokolova and Lapalme, 2009) based on the confusion matrix in Table 1 for the first class of disease and pest corn plant. For another class, the measures are similar.

$$\text{Accuracy} = \frac{\sum_{j=1}^{4} \frac{TP_j + TN_j}{TP_j + FP_j + FN_j TN}}{4} \tag{17}$$

$$\text{Precision} = \frac{\sum_{j=1}^{4} \frac{TP_j}{TP_j + FP_j}}{4} \tag{18}$$

$$\text{Recall} = \frac{\sum_{j=1}^{4} \frac{TP_j}{TP_j + FN_j}}{4} \tag{19}$$

$$F_1\text{Score} = \frac{2\text{Precision}(\text{Recall})}{(\text{Precision} + \text{Recall})} \tag{20}$$

$$\text{FPR} = 1 - \frac{\sum_{j=1}^{4} \frac{TN_j}{TP_j + FP_j}}{4} \tag{21}$$

A receiver operating characteristic (ROC) curve for a given model shows the trade-off between the recall and the false positive rate (FPR). FPR is the negation of specificity (TNR).

3. RESULTS AND DISCUSSION

Each digital image of 761 data on diseases and pests of corn plants is cropped and transformed into an RGB color space model with 32 x 32 pixels. The following is an example of a 32 x 32 matrix for each channel (component) R, G, and B from one of the Locusta Pest (LP) digital images,

$$\begin{bmatrix} 92 & 84 & 86 & \ldots & 115 \\ 94 & 96 & 94 & \ldots & 130 \\ 105 & 97 & 92 & \ldots & 130 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 73 & 108 & 100 & \ldots & 101 \end{bmatrix}$$

$$\begin{bmatrix} 101 & 94 & 94 & \ldots & 128 \\ 100 & 104 & 100 & \ldots & 142 \\ 108 & 100 & 97 & \ldots & 138 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 76 & 149 & 149 & \ldots & 153 \end{bmatrix}$$

$$\begin{bmatrix} 106 & 104 & 107 & \ldots & 137 \\ 112 & 115 & 114 & \ldots & 156 \\ 117 & 115 & 109 & \ldots & 151 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 91 & 45 & 45 & \ldots & 27 \end{bmatrix}$$

The pixel value of each component R, G, and B is the average value of all entries in the matrix 32x32.

The proposed validation model in this paper is a sub-sampling technique with a ratio of 70:30. Table 2 summarizes the composition of learning and testing data.

The Gaussian distribution assumption test, a requirement of the Naïve Bayes method, is presented in Table 3. The p-value for each component R, G, and B less than 5% in all tests indicates that all of these components are not Gaussian distribution.

Table 4 presents the discretizing components R, G, and B into five categories using equation (8). Each predictor variable has a different range of values in the same category.

This work also discretizes predictor variables into 2 and 3 categories using equation (8). Table 5 presents the predictive performance of the MNB method for the three discretizations. Based on Table 5, it can be seen that discretization into five categories has a lower performance measure than 2 and 3 categories. In both categories, all performance measures are the same, and only the AUC measure is different. AUC in 2 categories has a higher AUC. So, the MNB method with discretization into two categories has the highest prediction performance.

For the KNN method, we proposed a value of k=3,5,7,9,11 for the tuning system, where k is an odd number (because the number of classes in the dataset is even) starting from the smallest odd integer up to $\frac{1}{2}\sqrt{n}$ (with rounding), n =532 is the number of training data from the composition of 70: 30. Prediction performance using the KNN Method for several k is presented in Table 6. Based on Table 6, it can be seen that for k=3, the KNN method has the highest prediction performance.

Referring to Mishra et al. (2016), the performance of the KNN method is better than the performance of the MNB method. The performance of the MNB method in 2 categories was categorized as fair (AUC 70-80%), but the performance of the KNN method for all categories was categorized as excellent (AUC 90%). Likewise, when referring to (Karthik and Abhishek, 2019). The performance of the MNB method is categorized as good agreement (kappa 60-80%), but the performance of the KNN method for all k is categorized as very good (kappa 80%). Compared to Panigrahi et al. (2020), who also proposed

Table 5. Prediction Performance using the MNB Method in Percentage

Category	Accuracy	Precision	Recall	F_1-score	Kappa	AUC
2	92.72	79.88	79.24	78.17	72.44	71.91
3	92.72	79.88	79.24	78.17	72.44	71.76
5	88.94	69.75	64.76	66.81	58.06	64.07

Table 6. Prediction Performance using the KNN Method in Percentage

k	Accuracy	Precision	Recall	F_1-score	Kappa	AUC
3	98.54	88.57	94.38	93.59	94.3	95.45
5	98.11	87.57	93.4	92.51	92.57	93.94
7	97.96	85.04	92.72	91.49	92	92.42
9	97.67	84.16	91.32	90.64	90.85	92.42
11	97.53	84.03	90.68	90.21	90.28	92.42

the KNN and MNB methods to identify corn plant disease, the result of this work is better. As shown in Panigrahi et al. (2020), the performance measures for KNN are accuracy 76.16%, recall 75.00%, F_1-score 76.00%, and NB are accuracy 77.46%, recall 78.00%, F_1-score 75.50%.

4. CONCLUSIONS

Statistical machine learning has emerged as a critical component of modern scientific practice. Its presence enables automated techniques for predicting events, diagnosing phenomena, identifying objects based on prior observations, revealing hidden patterns in the data, and providing insight into the problem. Statistical machine learning methods have also been implemented for identifying corn plant diseases. In this work, we implement Multinomial Naïve Bayes and K-Nearest Neighbor methods to classify corn plant disease and pests. The results of our study have encouraging performance, especially the KNN method. The implementation of this method has a performance measure of the accuracy of 99.54%, a precision of 88.57%, recall 94.38%, F_1-score 93.59%, kappa 94.30%, and AUC 95.45%. For the MNB method, the performance measures are accuracy of 92.72%, a precision of 79.88%, recall 79.24%, F_1-score 78.17%, kappa 72.440%, and AUC 71.91%. These performance measures indicate the successful identification of corn plant diseases and pests. However, this success depends on the amount and quality of available data and the used statistical machine learning methods. The future work of this research is, first, collecting more quality data. Second, implement more other statistical machine learning methods. These works are expected to provide complete information about the best methods for identifying diseases and pests of the corn plant.

5. ACKNOWLEDGEMENT

The publication of this paper was supported by DIPA of Universitas Sriwijaya 2021 Public Service Agency, SP DIPA-23.17.2. 677515/2021, on November 23, 2021. On April 28, 2021, the Rector's Decree 0010/ UN9/ SK.LP2M. PT/2021 was issued.

REFERENCES

Adefisoye, J., G. Kibria, and F. George (2016). Performances of Several Univariate Tests of Normality: An Empirical Study. *Journal of Biometrics and Biostatistics*, **7**(4); 1–8

Alsafy, B. M., Z. M. Aydam, and W. K. Mutlag (2014). Multiclass Classification Methods: A Review. *International Journal of Advanced Engeineering Technology And Innovative Science*, **5**(3); 1–10

Arnastauskaitė, J., T. Ruzgas, and M. Bražėnas (2021). An Exhaustive Power Comparison of Normality Tests. *Mathematics*, **9**(788); 1–20

Chen, H. and D. Fu (2018). An Improved Naive Bayes Classifier for Large Scale Text. *Advances in Intelligent Systems Research*, **146**; 33–36

Dinesh, S. and T. Dash (2016). Reliable Evaluation of Neural Network for Multiclass Classification of Real-world Data. *ArXiv Rreprint*

Han, J., J. Pei, and M. Kamber (2011). *Data mining: concepts and techniques*. 225Wyman Street: Morgan Kaufmann Publishers

Hsu, S.-C., I.-C. Chen, and C.-L. Huang (2017). Image Classification Using Naive Bayes Classifier With Pairwise Local Observations. *Journal of Information Science & Engineering*, **33**(5); 1177–1193

Jäntschi, L. and S. D. Bolboacă (2018). Computation of Probability Associated with Anderson–Darling Statistic Lorentz. *Mathematics*, **6**(6); 88

Karthik, R. and S. Abhishek (2019). Machine Learning Using R: With Time Series and Industry-Based Use Cases in R. *Apress*, **2**(321); 1

Kresnawati, E. S., Y. Resti, B. Suprihatin, M. R. Kurniawan, and W. A. Amanda (2021). Coronary Artery Disease Prediction Using Decision Trees and Multinomial Naïve Bayes with k-Fold Cross Validation. *Inomatika*, **3**(2); 174–189

Kusumo, B. S., A. Heryana, O. Mahendra, and H. F. Pardede (2019). Machine Learning-based for Automatic Detection of Corn-Plant Diseases Using Image Processing. In *International Conference on Computer, Control, Informatics and Its Applications: Recent Challenges in Machine Learning for Computing Applications, IC3INA 2018 - Proceeding*; 93–97

Mengistu, A. D., S. G. Mengistu, and D. Alemayehu (2018). An Automatic Coffee Plant Diseases Identification using Hybrid Approaches of Image Processing and Decision. *Journal of Electrical Engineering and Computer Science*, **9**(3); 806–811

Mishra, S., O. A. Vanli, F. W. Huffer, and S. Jung (2016). Regularized Discriminant Analysis for Multi-Sensor Decision Fusion and Damage Detection with Lamb Waves Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems. *International Society for Optics and Photonics*, **9803**; 98032

Ngugi, L. C., M. Abelwahab, and M. Abo-Zahhad (2021). Recent Advances in Image Processing Techniques for Automated Leaf Pest an Diseas Recognition - A Review. *Information Processing in Agriculture*, **8**(1); 27–51

Pan, Y., H. Gao, H. Lin, Z. Liu, L. Tang, and S. Li (2018). Identification of Bacteriophage Virion Proteins using Multinomial Naïve Bayes with G-Gap Feature Tree. *International Journal of Molecular Sciences*, **19**(6); 1779

Panigrahi, K. P., H. Das, A. K. Sahoo, and S. C. Moharana (2020). Grape Leaf Disease Detection and Classification Using Machine Learning. In *n Progress in Computing, Analytics and Networking Proceedings of ICCAN 2019*; 659–669

Razali, N. M., Y. B. Wah, et al. (2011). Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *Journal of Statistical Modeling and Analytics*, **2**(1); 21–33

Resti, Y., E. S. Kresnawati, N. R. Dewi, N. Eliyati, et al. (2021). Diagnosis of Diabetes Mellitus in Women of Reproductive Age using the Prediction Methods of Naive Bayes, Discriminant Analysis, and Logistic Regression. *Science and Technology Indonesia*, **6**(2); 96–104

Rukmawan, S., F. Aszhari, Z. Rustam, and J. Pandelaki (2021). Cerebral Infarction Classification Using the K-Nearest Neighbor and Naive Bayes Classifier. *Journal of Physics: Conference Series*, **1752**; 012045

Sibiya, M. and M. Sumbwanyambe (2019). A Computational Procedure for the Recognition and Classification of Maize Leaf Diseases Out of Healthy Leaves Using Convolutional Neural Networks. *Agri Engineering*, **1**(1); 119–131

Sokolova, M. and G. Lapalme (2009). A Systematic Analysis of Performance Measures for Classification Tasks. *Information Processing and Management*, **45**(4); 427–437

Srianto, D. and E. Mulyanto (2016). Perbandingan K-Nearest Neighbor Dan Naive Bayes Untuk Klasifikasi Tanah Layak Tanam Pohon Jati. *Technology and Communication*, **15**(6); 241–245

Syarief, M. and W. Setiawan (2020). Convolutional Neural Network for Maize Leaf Disease Image Classification. *Telecommunication Computing Electronics and Control*, **18**(3); 1376–1381

Umar, R., I. Riadi, D. A. Faroek, et al. (2020). A Komparasi Image Matching Menggunakan Metode K-Nearest Neightbor (KNN) dan Support Vector Machine (SVM). *Journal of Applied Informatics and Computing*, **4**(2); 124–131

Xian, T. S. and R. Ngadiran (2021). Plant Diseases Classification using Machine Learning. *Journal of Physics: Conference Series*, **1962**(1); 012024

05_JURNAL_2022_Identification of Corn Plant Diseases and Pests Based on Digital Images using Multinomial Naïve Bayes and K-Nearest Neighbor

ORIGINALITY REPORT

11%

SIMILARITY INDEX

PRIMARY SOURCES

1	eprints.uad.ac.id Internet	51 words — 1%
2	www.omicsonline.org Internet	49 words — 1%
3	Lizhong Jiang, Kang Peng, Jian Yu, Wangbao Zhou, Yongjian Zuo. "The influence of trains on the seismic response of simply-supported beam bridges with different pier heights expressed through an impact factor", Bulletin of Earthquake Engineering, 2022 Crossref	30 words — 1%
4	www.bi.cs.titech.ac.jp Internet	28 words — 1%
5	D Adriani, Maryadi, E Purbiyanti, E Mulyana, A Minha. "Marginal Substitution and Allocative Efficiency Evaluation of Rice Farming Applying Organic Fertilizer In Swampy Land Banyuasin Regency, South Sumatera Province, Indonesia", IOP Conference Series: Earth and Environmental Science, 2022 Crossref	21 words — 1%
6	www.tandfonline.com Internet	21 words — 1%

7	Darmawi, , Riman Sipahutar, Siti Masreah Bernas, and Momon Sodik Imanuddin. "Renewable energy and hydropower utilization tendency worldwide", Renewable and Sustainable Energy Reviews, 2013. Crossref	15 words — < 1%
8	files.eric.ed.gov Internet	15 words — < 1%
9	Yanyuan Pan, Hui Gao, Hao Lin, Zhen Liu, Lixia Tang, Songtao Li. "Identification of Bacteriophage Virion Proteins Using Multinomial Naïve Bayes with g-Gap Feature Tree", International Journal of Molecular Sciences, 2018 Crossref	14 words — < 1%
10	Nova Yuliasari, Arini Badri, Patimah Mega Syah Bahar Siregar, Neza Palapa et al. "Improving the Performance of Mg/Cr LDH by Forming Metal Oxides Mg/Cr-Ni Using Coprecipitation Method as Adsorbent for Cationic Dyes", Journal of Ecological Engineering, 2022 Crossref	13 words — < 1%
11	inomatika.unmuhbabel.ac.id Internet	13 words — < 1%
12	"Loss Distribution Approaches", Measuring Operational and Reputational Risk, 2015. Crossref	12 words — < 1%
13	88836955-4e78-4855-9c8a-e3452c55ca11.filesusr.com Internet	12 words — < 1%
14	epdf.tips Internet	12 words — < 1%

15 Stavroyiannis, Stavros, Ilias Makris, Vasilis Nikolaidis, and Leonidas Zarangas. "Value-at-risk for the long and short trading position with the Pearson type-IV distribution", Global Business and Economics Review, 2013. — 11 words — < 1%
Crossref

16 Liu, Teng. "Nonparametric Tests for Informative Selection and Small Area Estimation for Reconciling Survey Estimates.", Colorado State University, 2020 — 10 words — < 1%
ProQuest

17 "Implementing a Self-measured Blood Pressure Monitoring Process", University of St. Augustine for Health Sciences Library, 2023 — 9 words — < 1%
Crossref Posted Content

18 "Proceeding of the 1st International Conference on Tropical Agriculture", Springer Science and Business Media LLC, 2017 — 9 words — < 1%
Crossref

19 Hyunsoo Kim, Changbum R. Ahn, Kanghyeok Yang. "Identifying Safety Hazards Using Collective Bodily Responses of Workers", Journal of Construction Engineering and Management, 2017 — 9 words — < 1%
Crossref

20 Mohammad Zulkarnain, Rico Januar Sitorus, Yeni Anna Appulembang, Annisah Biancika Jasmine, Abubakar Lutfi. "The Relationship between Stunting and Tooth Eruption of Primary School Children at Tuah Negeri Sub-district, Musi Rawas District, South Sumatera Province, Indonesia", Open Access Macedonian Journal of Medical Sciences, 2022 — 9 words — < 1%
Crossref

21 Sonam Pandey, Rohit Miri, G R Sinha, Rohit Raja. " AFD filter and E N classifier for improving — 9 words — < 1%

visualization of crop image and crop classification in remote sensing image ", International Journal of Remote Sensing, 2022
Crossref

22 pubs.rsc.org 9 words — < 1%
Internet

23 www.journalijar.com 9 words — < 1%
Internet

24 Jinlai Zhang, Yanmei Meng, Jin Wei, Jie Chen, Johnny Qin. "A Novel Hybrid Deep Learning Model for Sugar Price Forecasting Based on Time Series Decomposition", Mathematical Problems in Engineering, 2021 8 words — < 1%
Crossref

25 Karthik Ramamurthy, Rashmi Dinesh Thekkath, Shivam Batra, Sreejan Chattopadhyay. "A novel deep learning architecture for disease classification in Arabica coffee plants", Concurrency and Computation: Practice and Experience, 2023 8 words — < 1%
Crossref

26 Novitasari Ambarita, Restu Juniah, Syarifudin Syarifudin, Syaifudin Zakir, Hisni Rahmi. "Assessment on Environmental Quality Management of Coal Stockpile for Environmental Sustainability", Research Square Platform LLC, 2021 8 words — < 1%
Crossref Posted Content

27 link.springer.com 8 words — < 1%
Internet

28 www.insightsociety.org 8 words — < 1%
Internet

29 medium.com
Internet

7 words — < 1%

30 Berna Yazici, Senay Yolacan. "A comparison of various tests of normality", Journal of Statistical Computation and Simulation, 2007
Crossref

6 words — < 1%

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