J-INTECH (Journal of Information and Technology)
Accredited Sinta 4 Ministry of Higher Education, Science and Technology
Republic of Indonesia SK No. 10/C/C3/DT.05.00/2025
E-ISSN: 2580-720X || P-ISSN: 2303-1425

Identification of Palm Oil Fresh Fruit Bunches Worth Selling with K-Nearest Neighbors Algorithm

Dechy Deswita Indriani S[1], Kana Saputra S[2], Said Iskandar Al Idrus[3], Susiana[4], Adidtya Perdana[5]

[1,2,3,5]*Medan State University, Faculty of Mathematics and Natural Sciences, Computer Science, Jl. William Iskandar Ps. V Medan Estate, Indonesia*
[4]*Medan State University, Faculty of Mathematics and Natural Sciences, Mathematics, Jl. William Iskandar Ps. V Medan Estate, Indonesia*

Keywords
Classification; Fresh Fruit Bunches; KNN; Merchantability; Oil Palm

****Corresponding Author:***
dechydeswitaindriani@*mhs.unimed.ac.id*

Abstract
Indonesia is the world's largest palm oil producer, with annual production reaching more than 45 million tons. The quality of oil palm fresh fruit bunches (FFB) determines the quality of the oil produced. The quality of FFBs can be seen through their maturity and health. Fruit that is not ripe, overripe, or contaminated with mold can reduce oil quality due to high levels of free fatty acids (FFA). This research aims to build a classification model of FFB marketability using the K-Nearest Neighbors (K-NN) algorithm with RGB and GLCM features. Image data was collected from the plantation, then processed through the stages of preprocessing, feature extraction, and normalization. The model was tested in three approaches, namely using RGB-GLCM combination features, RGB only, and GLCM only, with various data sharing scenarios, namely 70:30, 80:20, and 90:10, as well as varying k values, namely k = 3, 5, 7, 9. The evaluation results show that the RGB-GLCM feature combination model in the 80:20 data sharing scenario and k = 5 value is the most optimal model, with accuracy reaching 88%. In addition to providing high accuracy, this model also shows good stability compared to the RGB and GLCM models alone. This proves that the use of a combination of features is more effective and reliable in identifying the marketability of oil palm FFB compared to the use of a single feature.

1. Introduction

Indonesia is one of the countries that has great potential in the agricultural sector and also plantations because it has fertile land that can support the agricultural sector. Palm oil is one type of plant that is used as a plantation commodity that has the opportunity to improve the economy in Indonesia [1]. However, despite its large contribution to the Indonesian economy, the main challenge for the country's palm oil industry lies in maintaining consistent product quality, especially to meet international standards. Based on data from the Central Bureau of Statistics, Indonesia is the largest palm oil producer in the world, with average production in 2018-2022 reaching more than 45 million tons per year [2].

Figure 1. Statistics on oil palm production in Indonesia [2]

Palm oil production is the result of extraction from fresh fruit bunches of oil palms. To produce good palm oil, the fresh fruit bunches must be of good quality. The quality of oil palm fresh fruit bunches is determined by several factors, especially the maturity and health of the fruit [3]. There are several maturity levels of oil palm FFBs, namely unripe, ripe, and overripe or almost rotten. The discoloration of the skin of the oil palm fruit and the number of small fruits that are detached are references to maturity. Unripe fruits have a black color and no loose fruits. Ripe fruit has a reddish orange color and there are 1 to 10 loose fruits. Overripe or almost rotten fruit has a reddish orange color and more than 50 detached loose fruits [4], [5]. In addition to maturity level, fruit health is also a factor in determining oil quality. Fruit health refers to morphological damage to oil palm fresh fruit bunches caused by microorganisms. What causes the fruit to be contaminated by microorganisms is because it is placed in a dirty and humid environment, fruit damage can also occur during the harvesting process [6].

The process of sorting or selecting oil palm FFBs is carried out manually by sorting officers (graders) both at the collection point and the palm oil mill. This process is done to ensure that the selected oil palm FFBs meet the right maturity and health standards. However, this manual assessment tends to be subjective and varies between graders, which can lead to inconsistencies in determining the quality of oil palm FFB. Thus, palm oil FFBs that do not meet the standards sometimes pass to the production stage, and then affect the quality and stability of the final palm oil product produced [7], [8]. Inaccuracies in determining the quality of palm oil FFB can reduce the economic value of the oil produced and cause losses to producers and also have a negative impact on the image of the Indonesian palm oil industry in the global market [9]. One potential solution that can be applied is to utilize machine learning, which is a branch of artificial intelligence that allows computers to learn from data and make decisions or predictions without the need to be explicitly programmed [10].

K-Nearest Neighbors (K-NN) is a simple yet highly effective machine learning algorithm for classification. It works by comparing the data to be classified with existing training data based on a certain distance, such as Euclidean distance. The new object will be classified into the class that appears the most among its nearest neighbors (K). This method has the advantage of handling datasets with not too high dimensions and ease of implementation and interpretation [11]. This algorithm has been widely implemented by previous researchers. In [12], the K-NN algorithm was applied in analyzing the sentiment of twitter users on issues related to government policies on online learning, and obtained an accuracy of 84.56% when K = 10. Research from [1] also used the K-NN algorithm in classifying the maturity level of guava based on digital image processing, and obtained an accuracy rate of 93% with K = 1. Comparison of the K-NN algorithm with SVM (Support Vector Machine) was also carried out by research [13] in predicting the secondary structure of proteins. The results obtained confirmed that the presence of certain amino acids in certain protein sequences increases stability for protein secondary structure prediction, and the K-NN algorithm has better performance in predicting protein secondary structure compared to the SVM algorithm. In study [14], the Linear Discriminant Analysis (LDA) algorithm with Mahalanobis distance and K-Nearest Neighbors (K-NN) were used to classify the quality of fresh fruit bunches (FFB) of oil palm based on optical sensor data. The model evaluation results showed that the K-NN algorithm achieved the best accuracy compared to the LDA algorithm based on grader assessment. The accuracy was 80.7% for K-NN and 79.8% for LDA. Meanwhile, the accuracy based on TBS oil content reached 88.2% for both algorithms.

In this research, feature extraction becomes an important part of the analysis process, as image data is complex and requires a simpler and still informative representation[15]. The two types of features extracted are color features using the RGB (Red, Green, Blue) model and texture features using the GLCM (Gray Level Co-occurrence Matrix) method. The selection of RGB color and GLCM texture features has strong reasons in the context of this research. Color represents the ripeness of the fruit, which is highly relevant in determining the marketability of oil palm FFBs, while texture provides additional information about the health of the fruit. By combining these two features, the model can evaluate the condition of oil palm FFB holistically, covering two main aspects, namely fruit maturity and health. So based on this background, the title “Identification of Palm Oil Fresh Fruit Bunches Worth Selling with K-Nearest Neighbors Algorithm” was raised.

2. Research Method

This research was conducted at the collection point of a private oil palm plantation located in Dusun Batang Kandis KM. 90, Kandis Village, Kandis District, Siak Regency, Riau, for 28 days, namely March 01 - 31, 2025. The type of research used was experimental, which aims to test the effect of color and texture features on the classification of the marketability of oil palm FFB. The research population was all FFBs that have been harvested at the TPH, with a sample of 200 image data evenly divided into two classes: marketable and unmarketable. This relatively small sample size is one of the limitations of the study because it can affect the model's generalization ability. However, the use of the K-Nearest Neighbors (KNN) algorithm is considered appropriate because it is a lazy learner, efficient on limited datasets, and capable of producing stable performance if the data has undergone pre-processing and normalization [16],[17]. The independent variables in this study are color (RGB) and texture features (GLCM: contrast, correlation, energy, homogeneity), while the dependent variable is the result of the classification of FFB marketability based on the level of maturity and morphological condition of the fruit. Primary data in the form of FFB images were collected using a high-resolution DSLR camera. Validation was conducted by comparing camera images and direct observation by palm oil experts, as well as reliability testing through repeated image capture. The research procedure includes the stages of data collection, pre-processing (background removal, resize, cropping), feature extraction (RGB and GLCM), and the construction of a classification model using the K-Nearest Neighbors (KNN) algorithm with the k value parameter and Euclidean distance measurement. The extracted features are normalized and divided into training data and test data with 70:30, 80:20, and 90:10 division scenarios. The extracted features are normalized and divided into training data and test data with a division scenario of 70:30, 80:20, and 90:10. The resulting model is then tested and evaluated using confusion matrix (accuracy, precision, recall, f1-score) to assess the performance of the classification of oil palm FFB salability.

3. Result and Discussions

At this stage, the original background of each image data is removed and converted to white [18].

Figure 2. Image before and after background removal

The image data is cropped to remove irrelevant parts, so as to focus more on the main object.

Figure 3. Image before and after cropping

The image dimensions are changed to be uniform, to ensure consistency in the dataset [18]. The image is resized evenly to 500x500 pixels.

Figure 4. Image before and after resizing

The feature extraction stage is carried out to identify and extract special features that characterize the image. The extraction features used to identify images in this study are RGB (Red, Green, Blue) and GLCM (Gray Level Co-occurrence Matrix).

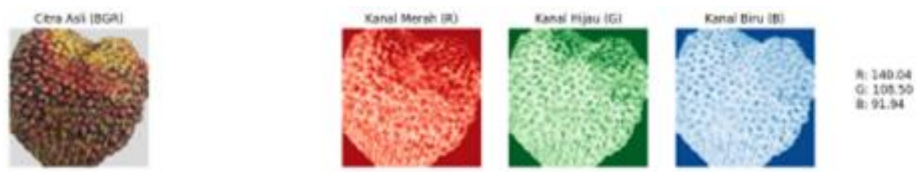

Figure 5. RGB feature extraction

Figure 5 above is one example of palm oil FFB image data that has been extracted RGB features, displaying the original image converted into each RGB color channel along with the value of the extraction results of each color (R, G, B).

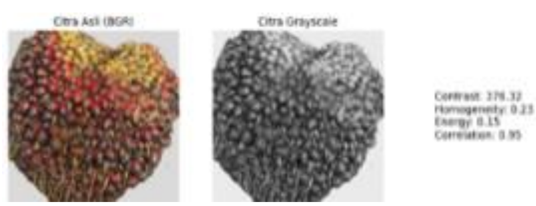

Figure 6. GLCM feature extraction

Figure 6 above is one example of palm oil FFB image data that has been extracted GLCM features, displaying the original image converted into grayscale form, along with the value of the extraction results of each feature (contrast, homogeneity, energy, correlation).

]The value scale between features in this research dataset is not uniform, so normalization is carried out to equalize the scale, so that there are no features that dominate by reason of having a larger value scale[19]. The normalization method used in this research is Min-Max Normalization, by changing the scale of each feature into the value range [0, 1], using the equation (1):

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Where, x is data to be normalized, x_{norm} is normalized data, x_{min} is minimum data in features, and x_{max} is maximum data in features.

Tabel 1. Dataset after normalization

No	Label	R	G	B	Contrast	Homogeneity	Energy	Correlation
1	L	0.6358	0.4935	0.2670	0.232	0.19318	0.1899	0.3491
2	L	0.6330	0.6581	0.4843	0.4417	0.2020	0.2569	0.1038
3	L	0.7772	0.7471	0.6118	0.2852	0.3507	0.4115	0.3914
4	L	0.8603	0.8031	0.4924	0.2530	0.3170	0.3445	0.4579
5	L	0.6918	0.5823	0.4339	0.5127	0.2103	0.2688	0.0817
...		....	....	....	....	....	....	....
196	TL	0.3104	0.4973	0.5666	0.0605	0.3123	0.2470	0.8594
197	TL	0.4528	0.7441	0.7631	0.0829	0.5961	0.5883	0.9137
198	TL	0.2924	0.5156	0.5995	0.0914	0.5050	0.4105	0.8983
199	TL	0.3257	0.5412	0.6007	0.1970	0.3732	0.3587	0.7815
200	TL	0.5067	0.5851	0.5490	1	0.2214	0.2850	0.0840

In Tabel 1, the L label refers to the Appropriate class and the TL label refers to the Inappropriate class.

Before entering the classification process, the dataset will be divided into train data and test data first (splitting data). In this research, the dataset division uses three scenarios, namely 70:30, 80:20, and 90:10 for all models built.

Tabel 2. Dataset division

Description	70:30	80:20	90:10
Data x train	140	160	180
Data x test	60	40	20
Data y train	140	160	180
Data y test	60	40	20

After various data pre-processing, the next step is to build a classification model. This classification model will be built using the K-Nearest Neighbors (KNN) algorithm. KNN is an algorithm that groups objects based on high similarity with other objects, by paying attention to the smallest distance. To build the KNN model, the initial parameters, namely the K value must first be determined. The distance calculation used is the Euclidean Distance method, which will be contained in equation (2).

$$d_{(x,y)} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (2)$$

Where,

$d_{(x,y)}$ = Distance between two points x and y

x_i = Feature value *i* of point x

y_i = Feature value *i* of point y

n = Number of features

In this research, four scenarios of k value will be used, namely k = 3, 5, 7, and 9. The selection of k values was done manually, considering that k values that are too small can cause overfitting, while k values that are too large can cause underfitting. Odd values were chosen to avoid tied or ambiguous voting results[20]. The best k value will then be selected for each model.

The test example will use 10 test data that has been previously divided, the data used will be contained in Tabel 3 below.

Tabel 3. Training data

ID	R	G	B	contras	homogeneity	energy	correlation	Label
A	0.3018	0.1749	0.1908	0.3482	0.3457	0.2442	0.7305	L
B	0.4597	0.6479	0.6323	0.8159	0.2220	0.3018	0.2818	TL
C	0.7177	0.5554	0.5666	0.5131	0.5979	0.6886	0.3837	L
D	0.5060	0.4581	0.2820	0.6314	0.2239	0.2198	0.4928	TL
E	0.3830	0.4739	0.4250	0.2995	0.6539	0.6395	0.8333	TL
F	0.4151	0.5476	0.5338	0.4444	0.5542	0.5769	0.7203	TL
G	0.8334	0.9001	0.7382	0.7931	0.6720	0.7807	0.4928	L
H	0.3862	0.3458	0.3724	0.3420	0.4131	0.3888	0.7350	L
I	0.6178	0.3835	0.4115	0.1389	0.7706	0.6933	0.8243	L
J	0.4336	0.4771	0.4391	0.9043	0.0362	0.1015	0.0998	TL

The test data in this test example will be contained in Tabel 4 below:

Tabel 4. Example

ID	R	G	B	contras	homogeneity	energy	correlation	Label
N	0.3018	0.1749	0.1908	0.3482	0.3457	0.2442	0.7305	**?**

Next, we will calculate the distance between test data and training data using the Euclidean Distance method in Tabel 5 below:

Tabel 5. Euclidean distance of each data

No	Describe	Euclidean Distance	Label
1	d(N,A)	0,662411	L
2	d(N,B)	0,907214	TL
3	d(N,C)	0,538132	L
4	d(N,D)	0,751377	TL
5	d(N,E)	0,198129	TL
6	d(N,F)	0,231925	TL
7	d(N,G)	0,790906	L
8	d(N,H)	0,382628	L
9	d(N,I)	0,348655	L
10	d(N,J)	1,211601	TL

After the distance of each data is known, we will sort the distance values from the smallest to the largest [7]. The sorting of distances will be contained in Tabel 6 below:

Tabel 6. Euclidean distance sorting

No.	Describe	Euclidean Distance	Label	Sorting
1	d(N,A)	0,662411	L	6
2	d(N,B)	0,907214	TL	9
3	d(N,C)	0,538132	L	5
4	d(N,D)	0,751377	TL	7
5	d(N,E)	0,198129	TL	1
6	d(N,F)	0,231925	TL	2
7	d(N,G)	0,790906	L	8
8	d(N,H)	0,382628	L	4
9	d(N,I)	0,348655	L	3
10	d(N,J)	1,211601	TL	10

The k value used for this test example is k = 5. Thus, the data that will be processed to the next stage is only the data that is in the order of 1 to 5 to determine the data label N based on the majority category voting [7].

Tabel 7. Determination of the majority category

Describe	Euclidean Distance	Label	Sorting
d(N,E)	0,198129	TL	1
d(N,F)	0,231925	TL	2
d(N,I)	0,348655	L	3
d(N,H)	0,382628	L	4
d(N,C)	0,538132	L	5

Based on Tabel 7 above, it can be seen that the TL (Inappropriate) category obtained 2 votes, and the L (Appropriate) category 3 votes, so the Appropriate category is superior. Therefore, data N belongs to the Appropriate category.

In this study, experiments were conducted on the distribution of data, namely 70:30, 80:20, and 90:10. Not only in data division, experiments were also conducted on the value of k, namely k = 3, k = 5, k = 7, and k = 9. Based on the experiments conducted, the following information was obtained:

Tabel 8. Comparison of feature combination model accuracy

Describe	70:30	80:20	90:10
k = 3	75%	75%	80%
k = 5	83%	88%	80%
k = 7	80%	82%	80%
k = 9	75%	78%	80%

The accuracy results of each scenario in the feature combination model are listed in Tabel 8 above. The best accuracy in the feature combination model is 88% with a data split of 80:20 and a value of k = 5.

Tabel 9. Comparison of RGB model accuracy

Describe	70:30	80:20	90:10
k = 3	72%	70%	85%
k = 5	73%	82%	95%
k = 7	73%	75%	90%
k = 9	80%	80%	90%

The accuracy results of each scenario in the RGB model are contained in Tabel 9 above. The best accuracy in the RGB model is 95% with a data split of 90:10 and a value of k = 5.

Tabel 10. Comparison of GLCM model accuracy

Describe	70:30	80:20	90:10
k = 3	68%	70%	75%
k = 5	75%	80%	70%
k = 7	73%	75%	70%
k = 9	72%	75%	70%

Based on the accuracy comparison information presented above, the highest accuracy of each model will be selected. In the feature combination model, at 80:20 data division with a value of k = 5, 88% accuracy is obtained. The RGB model, at a data division of 90:10 with a value of k = 5, obtained an accuracy of 95%. The GLCM model, at a data division of 80:20 with a value of k = 5, obtained an accuracy of 80%. These three models will then be further analyzed in the model evaluation stage to see which model has better and stable performance, without any indication of overfitting.

A comparison of the train accuracy and test accuracy values contained in Tabel 11 was conducted, to determine the model that has better performance and is also stable.

Tabel 11. Comparison of performance

Model	Splitting	Train Accuracy	Test Accuracy	Difference
Feature Combination	80:20	0.8187	0.875	0.0563
RGB	90:10	0.8333	0.95	0.1167
GLCM	80:20	0.7625	0.80	0.0375

The RGB model is a model that has a very high accuracy of 95%. Although the RGB-only model produced the highest accuracy (95%), the significant difference between train accuracy and test accuracy (11%) indicated overfitting, a condition where the model adapts too much to the training data, thereby reducing its ability to generalize to the test data. This occurs because color features (RGB) are sensitive and able to distinguish classes well in training data, but are less robust to variations in conditions such as lighting and image quality. In contrast, the model with a combination of features (RGB + GLCM), although producing lower accuracy (88%), shows better classification stability with a train-test accuracy difference of only 5%. This indicates that the model is better able to generalize to new data and does not rely solely on color features, but also considers texture information that strengthens the classification process. Thus, feature combination was chosen because it provides a balance between accuracy and generalization ability, making it more optimal for application in the identification of fresh palm oil fruit bunches suitable for sale.

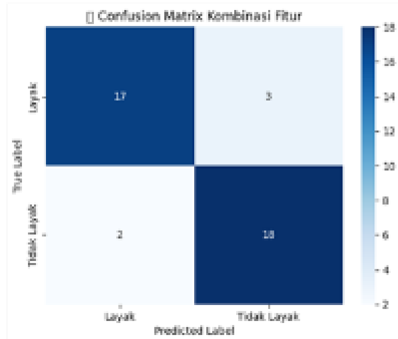

Figure 8. Confusion matrix of feature combination model

Based on Figure 8 which displays the confusion matrix heatmap of the feature combination model above, it can be seen that the model does not correctly predict 5 data out of 40 test data. Where, 3 data that have the original Eligible label are predicted to be Ineligible, and 2 data with the original Ineligible label are predicted to be Eligible. The feature combination model in the 80:20 data division scenario and k = 5 value obtained 88% accuracy, 88% precision, 88% recall, and 87% f1-score based on the model building report.

4. Conclusions and Future Works

Based on the results of the research, the development of a classification model for the salability of oil palm FFB using the K-Nearest Neighbors (KNN) algorithm was successfully carried out through the stages of collecting image data in the field, pre-processing (background removal, cropping, resizing), extracting color (RGB) and texture (GLCM) features, and normalizing with Min-Max Scaler. The KNN model was tested on various scenarios of k value and data division. The analysis results show that RGB features produce the highest accuracy (95%) in the 90:10 scenario with k = 5, but tend to overfitting. Meanwhile, the GLCM-based model showed the lowest performance with a maximum accuracy of 80%. The combination model of RGB and GLCM features proved to be the most optimal and stable, with 88% accuracy, 88% precision, 88% recall, and 87% f1-score in the 80:20 scenario with k = 5. The small difference in accuracy between training and test data (5.63%) indicates good model stability. Thus, the feature combination model is considered the most effective in classifying the salability of oil palm FFB. This research can be further developed by adding several indicators of palm fruit marketability, such as fruit size and fruit type, with a larger dataset, as well as implementation in website-based and Android applications for direct use in the field.

5. References

[1] S. Yuandry and I. Irdawati, "Analisis Pengaruh Tingkat Free Fatty Acid (FFA) terhadap Kualitas Crude Palm Oil (CPO) pada Pabrik Industri Kelapa Sawit PT. Agro Muko Po. Mill," *Biocelebes*, vol. 18, no. 1, pp. 13–19, 2024, doi: 10.22487/bioceb.v18i1.16882.

[2] BPS, "Statistik Kelapa Sawit Indonesia 2022," vol. 16, p. 6, 2023.

[3] I. Bonet, M. Gongora, F. Acevedo, and I. Ochoa, "Deep Learning Model to Predict the Ripeness of Oil Palm Fruit," *Int. Conf. Agents Artif. Intell.*, vol. 3, no. Icaart, pp. 1068–1075, 2024, doi: 10.5220/0012434600003636.

[4] M. Rifqi, M. AKbar, and Y. Fitrisia, "Aplikasi Pendeteksian Kematangan Tandan Buah Segar (TBS) Kelapa Sawit Berdasarkan Komposisi Warna Menggunakan Algoritma K-NN," *J. Komput. Terap.*, vol. 6, no. 1, pp. 99–107, 2020, doi: https://doi.org/10.35143/jkt.v6i1.3338.

[5] J. Zulkarnain, Kusrini, and T. Hidayat, "Klasifikasi Tingkat Kematangan Tandan Buah Segar Kelapa Sawit Menggunakan Pendekatan Deep Learning," *JST (Jurnal Sains dan Teknol.*, vol. 12, no. 3, pp. 748–758, 2024, doi: 10.23887/jstundiksha.v12i3.59140.

[6] D. Nurfiqih, L. Hakim, and M. Muhammad, "Pengaruh Suhu, Persentase Air, Dan Lama Penyimpanan Terhadap Persentase Kenaikan Asam Lemak Bebas (Alb) Pada Crude Palm Oil (Cpo)," *J. Teknol. Kim. Unimal*, vol. 10, no. 2, p. 1, 2021, doi: 10.29103/jtku.v10i2.4955.

[7] R. Kurniawan, Samsuryadi, F. S. Mohamad, H. O. L. Wijaya, and B. Santoso, "Classification of palm oil fruit ripeness based on AlexNet deep Convolutional Neural Network," *Sinergi (Indonesia)*, vol. 29, no. 1, pp.

207–220, 2025, doi: 10.22441/sinergi.2025.1.019.
[8] R. Malyala, "Development of a Convolutional Neural Network Model for Automated Ripeness Classification of Palm Oil Fresh Fruit Bunches," *Int. J. Innov. Sci. Res. Technol.*, vol. 9, no. 11, pp. 1040–1046, 2024, doi: 10.38124/ijisrt/ijisrt24nov762.
[9] M. Rifqi and Suharjito, "Deteksi Kematangan Tandan Buah Segar (Tbs) Kelapa Sawit Berdasarkan Komposisi Warna Menggunakan Deep Learning," *J. Tek. Inform. Atmaluhur*, vol. 6, no. 1, p. 40, 2021, doi: https://doi.org/10.15408/jti.v14i2.23295.
[10] L. Muflikhah, W. F. Mahmudy, and D. Kurnianingtyas, *Machine Learning*, 1st ed. Malang: UB Press, 2023. [Online]. Available: https://play.google.com/books/reader?id=tu_uEAAAQBAJ&pg=GBS.PR4&hl=id
[11] Rahmaddeni, D. Wulandari, M. Renova, G. Ramadhan, and R. Sari, *Machine Learning*, 1st ed. 2024. [Online]. Available: https://books.google.co.id/books?hl=id&lr=&id=owoOEQAAQBAJ&oi=fnd&pg=PA8&dq=buku+machine+learning&ots=X-lpMPhu2V&sig=mpfvmmnXntS5KJkuM8tG37MEogQ&redir_esc=y#v=onepage&q=buku machine learning&f=false
[12] A. R. Isnain, J. Supriyanto, and M. P. Kharisma, "Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 15, no. 2, p. 121, 2021, doi: 10.22146/ijccs.65176.
[13] A. Tasari, D. Dinata Tarigan, E. Nia, D. B. Purba, K. Saputra, and I. Artikel, "Perbandingan Algoritma Support Vector Machine dan KNN dalam Memprediksi Struktur Sekunder Protein," *J. Inform.*, vol. 9, no. 2, pp. 172–179, 2022, [Online]. Available: https://raw.githubusercontent.com/JiayingLi/Pr
[14] A. W. Setiawan, R. Mengko, A. P. H. Putri, D. Danudirdjo, and A. R. Ananda, "Classification of palm oil fresh fruit bunch using multiband optical sensors," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 4, pp. 2386–2393, 2019, doi: 10.11591/ijece.v9i4.pp2386-2393.
[15] R. Mahadyanto, D. A. Prastiningtyas, and F. E. Purwiantono, "Penerapan Metode Jaringan Syaraf Tiruan Radial Basis Function Untuk Identifikasi Jenis Mangga Berdasarkan Pola Daun," *J-INTECH (Journal Inf. Technol.*, vol. 07, no. 01, pp. 90–96, 2019, doi: https://doi.org/10.32664/j-intech.v7i01.411.
[16] Y. Dimas Pratama and A. Salam, "Comparison of Data Normalization Techniques on KNN Classification Performance for Pima Indians Diabetes Dataset," *J. Appl. Informatics Comput.*, vol. 9, no. 3, pp. 693–706, 2025, doi: 10.30871/jaic.v9i3.9353.
[17] M. S. Fajri, N. Septian, and E. Sanjaya, "Evaluasi Implementasi Algoritma Machine Learning K-Nearest Neighbors (kNN) pada Data Spektroskopi Gamma Resolusi Rendah," *Al-Fiziya J. Mater. Sci. Geophys. Instrum. Theor. Phys.*, vol. 3, no. 1, pp. 9–14, 2020, doi: 10.15408/fiziya.v3i1.16180.
[18] D. D. Indriani.S, E. J. A. Sinaga, G. Oktavia, H. Syahputra, and F. Ramadhani, "Identifikasi Tanda Tangan Dengan Menggunakan Metode Convolution Neural Network (CNN)," *J-INTECH (Journal Inf. Technol.*, vol. 12, no. 1, pp. 138–147, 2024, doi: 10.32664/j-intech.v12i1.1273.
[19] P. P. Allorerung, A. Erna, M. Bagussahrir, and S. Alam, "Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 9, no. 3, pp. 178–191, 2024, doi: 10.14421/jiska.2024.9.3.178-191.
[20] F. A. Tyas, M. Narayuni, and H. Rakhmawati, "Optimasi Algoritma K-Nearest Neighbors Berdasarkan Perbandingan Analisis Outlier (Berbasis Jarak, Kepadatan, LOF)," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 13, no. 02, pp. 108–115, 2024, doi: 10.22146/jnteti.v13i2.9579.