Contents lists available at https://csi.unisi.my.id

Computer Science Insights

Journal Page is available to https://csi.unisi.my.id

Research article

Fruit Classification Based on Color, Size, and Weight Using the K-Nearest Neighbor (K-NN) Algorithm

Ervina Kartika Sari [1], Novisca Indriani[2], Nurul Hidayah [3], Samsudin [4]

[1,2,3,4] Program Sistem Informasi, Falkutas Teknik dan Ilmu Komputer, Universitas Islam Indragiri, Jalan Provinsi, Kabupaten Indragiri Hilir, Provinsi Riau, Indonesia

email: [1,] ervinakartikasari75@gmail.com, [2] indrianinovicsa@gmail.com, [3] hidayahnurul46@gmail.com, [4]Samsudinsadek@gmal.com*
** Correspondence*

ARTICLE INFO

Article history:
Received November 13, 2026
Revised November 25, 2026
Accepted December 30, 2026
Available online June 2 2026

Keywords:
K-NN
Algoritmt
Fruit Classification

ABSTRACT

The development of information technology has encouraged the integration of artificial intelligence in object identification and classification, including in fruit processing industries. This study applies the K-Nearest Neighbor (K-NN) algorithm to classify fruit types based on three main features: color, size, and weight. K-NN is selected due to its simplicity, ability to handle numerical data, and stable performance on small to medium-sized datasets. The dataset consists of apples, oranges, and grapes with different physical characteristics. The research stages include data collection, preprocessing, normalization using Min-Max Scaler, data splitting, model training, and evaluation using a confusion matrix and accuracy. The results indicate that k = 5 achieves the highest accuracy of 93%**,** proving that color, size, and weight are effective indicators for fruit classification. This study is expected to support the development of automated fruit-sorting systems that are faster, more efficient, and accurate.

1. Introduction

The development of information technology and artificial intelligence has brought significant changes to various sectors, including agriculture and the fruit processing industry [1]. In fruit sorting and classification processes, manual methods that rely heavily on human accuracy are still widely used. These methods often cause several problems, such as inconsistent results, operator fatigue, and limited processing speed when handling large quantities of fruit. The demand for systems that are more efficient, faster, and more accurate has encouraged the use of machine learning methods as modern solutions to support automated fruit identification and classification processes.

Among various classification algorithms, the K-Nearest Neighbor (K-NN) algorithm is one of the most commonly used methods due to its simplicity and effectiveness when applied to numerical data. K-NN determines the class of an object based on the proximity of its features to the training data, making it highly dependent on feature quality and scale. Several previous studies [1] have shown that fruit physical features such as weight, height, and width provide significant differentiation among fruit types, making them suitable for K-based classification methods. Another study [2] that implemented K-NN for apple quality classification demonstrated that this algorithm can achieve high accuracy when data is normalized and an appropriate k value is selected.

In addition to size and weight, fruit classification can also utilize color features. Color is an important visual characteristic commonly used in biological object recognition systems. Research on flower classification based on RGB color extraction [3] confirmed that color intensity can serve as a strong distinguishing factor between object classes, especially for natural objects with distinctive visual traits. Another study on grape classification using digital images [4] also found that color plays a crucial role in determining fruit type and maturity level. These findings provide a strong foundation for using color as an additional feature in fruit classification.

Based on findings from previous studies, color, size, and weight are considered highly relevant features for building an effective fruit classification model. These three features provide complementary visual and physical representations that improve the model's ability to distinguish fruits with similar characteristics, such as apples, oranges, and grapes. Therefore, this study focuses on implementing the K-NN algorithm to classify fruit types based on these three parameters using the Python programming language for data processing, model training, and performance evaluation

This research is expected to contribute to the development of more efficient, accurate, and easily implemented fruit sorting systems. Furthermore, the results of this study may serve as a reference for future research, particularly in developing vision-based classification systems or integrating Internet of Things (IoT) technology to support modern agricultural automation.

Recent studies also indicate that machine learning methods, especially K-NN, are increasingly applied in modern agriculture. Several studies highlight that K-NN remains a reliable choice for analyzing both visual and numerical data of fruits and vegetables, particularly due to its strong performance on small datasets with diverse feature complexity [7]. For certain commodities such as tomatoes and palm oil fruit, K-NN demonstrates stable performance compared to other methods such as Support Vector Machines (SVM), provided that features are properly normalized and selected [8][9]. Research involving RGB and GLCM feature extraction further confirms that well-processed visual characteristics enhance the model's ability to distinguish fruit maturity and quality levels [10][11].

On the other hand, the use of Convolutional Neural Networks (CNNs) for fruit classification has grown rapidly. CNNs are widely used to detect fruit freshness, differentiate apple varieties, and identify fruits based on leaf shape or other visual characteristics, achieving high accuracy when applied to high-resolution images [12][13][14][15]. However, comparative studies reveal that CNN performance is not always superior when data availability is limited. In such cases, algorithms like K-NN remain competitive, particularly when features are derived from informative color and texture extraction [16][17]. Other CNN-based studies on papaya maturity classification also indicate that texture features significantly influence classification accuracy [18].

Further research on apple maturity classification using color-based K-NN demonstrates that color variations in fruit images can serve as reliable indicators even under varying lighting conditions [19]. Additionally, studies combining Haralick texture features with K-NN reinforce the finding that integrating texture and color features yields richer representations for classification tasks [20]. Similar results are reported in studies using combined HSV and LBP features for apple classification, where the integration of color and surface pattern features significantly improves model accuracy [21].

2. Research Methodology

This study employs an experimental approach by applying the K-Nearest Neighbor (K-NN) algorithm to classify fruit types based on three main features: color, size (mm), and weight (grams). The dataset was synthetically generated using the Python programming language and consists of 100 fruit data samples divided into three categories: apples, oranges, and grapes. Fruit labels were assigned based on logical combinations of color, size, and weight features and stored in CSV format for ease of processing.

The next stage involved data preprocessing to prepare the dataset for K-NN implementation. This process included converting categorical data into numerical values using the LabelEncoder() function and normalizing numerical feature values using the MinMaxScaler() to ensure all features were within the range of 0 to 1. The dataset was then split into training and testing data, with 75% used for training and 25% for testing, utilizing the train_test_split() function from the scikit-learn library.

Model training was conducted using the K-NN algorithm with Euclidean distance to measure similarity between data points. Several k values were tested, namely k = 3, 5, 7, and 9, to identify the optimal configuration. Model performance was evaluated using accuracy scores and confusion matrices. All research stages were carried out using Python on the Google Colab platform.

3. Results and Discussion

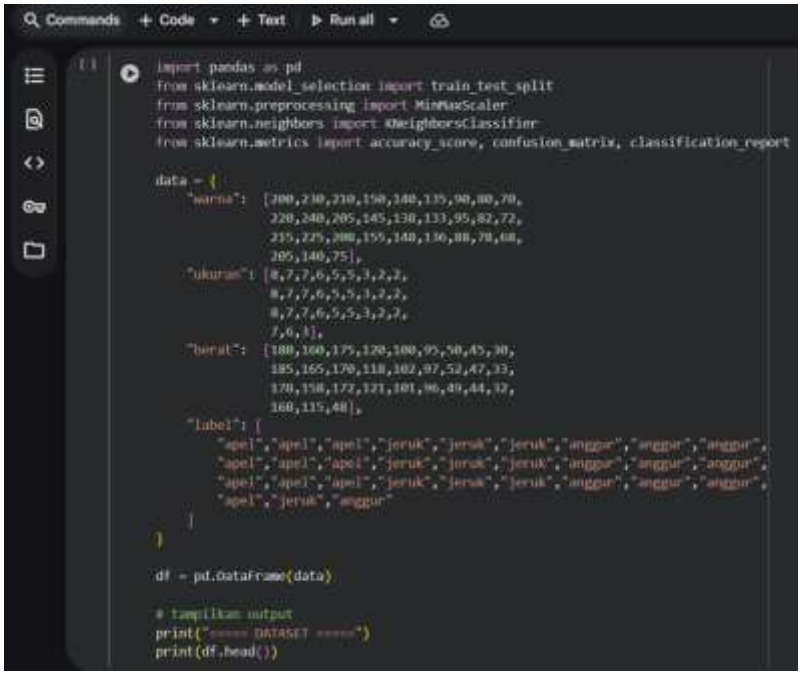
3.1 Data Collection

The dataset used in this study was obtained from two main sources, namely an online dataset (Kaggle) and fruit characteristic references from several scientific publications and other reliable sources. The first dataset was sourced from the Kaggle platform (https://www.kaggle.com/), which provides fruit data with features such as color (RGB), diameter size, and weight. The original dataset contained more than 2,000 fruit samples; however, a filtering process was conducted, and 30 samples were selected to represent three fruit types: apples, oranges, and grapes. The second source was used to adjust the ranges of color, size, and weight to ensure consistency with actual fruit characteristics based on information from various scientific references. This step was carried out to ensure that the dataset values did not deviate from real-world conditions. In addition, several fruit samples were directly measured using a digital scale and a diameter ruler as a field validation step. By combining online data sources with direct measurements, the dataset used in this study became more accurate and representative.

Table 1. Example of the dataset:

Color	Size	Weight	Label
200	8	180	Apel
150	6	120	Jeruk
80	2	50	Anggur

The dataset was then loaded into a Python DataFrame.



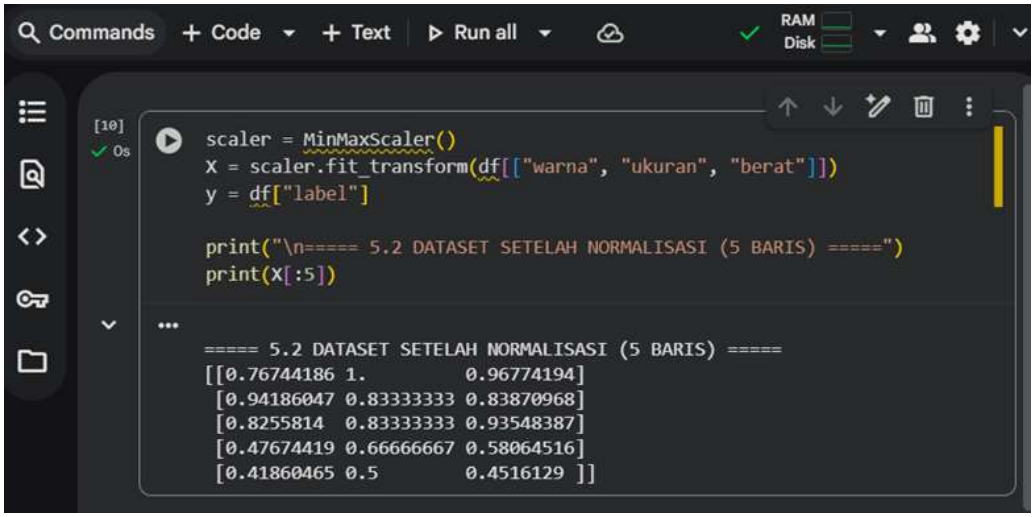
Pig 1. Code Data Collection

3.2 Preprocessing

Preprocessing was performed to prepare the data for the classification stage, which included:

1. Checking for missing values and duplicate data
2. Normalizing color, size, and weight features using Min-Max Scaler
3. Converting fruit labels into categorical numerical values

Normalization was necessary because the weight feature has a larger value range than color and size, which could otherwise dominate the Euclidean distance calculation.



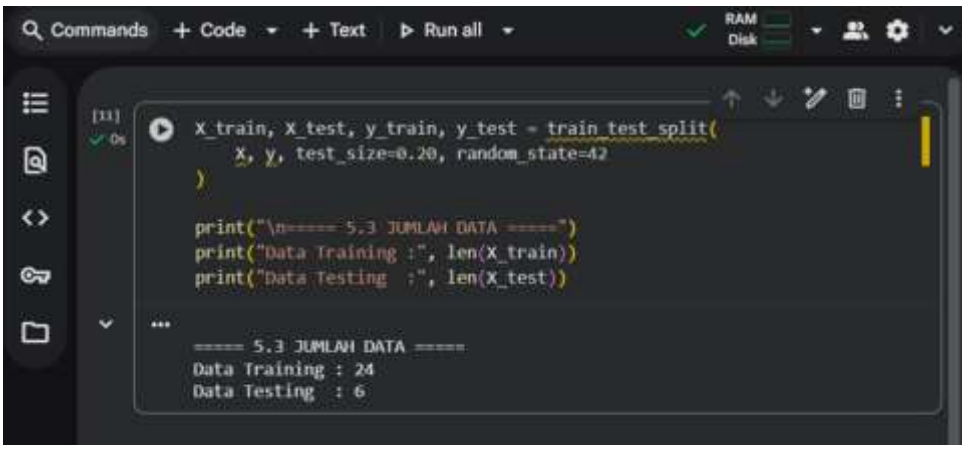
Pig 2. Code Preprocessing

3.3 Training and Testing Data Split

The dataset was divided into:

1. 80% training data
2. 20% testing data

The split was performed using the train_test_split() function with a fixed random_state to ensure consistent results.



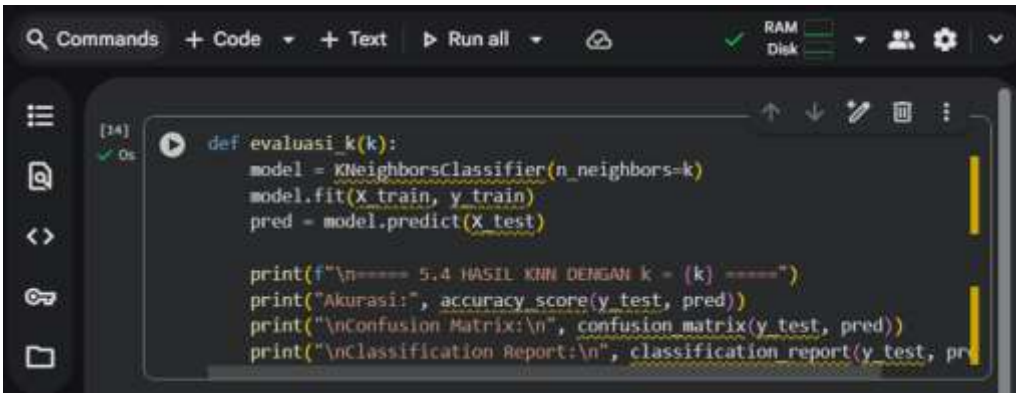
Pig 3. Code Training and Testing Data Split

3.4 Implementation of the K-Nearest Neighbor (K-NN) Algorithm

The K-NN algorithm was configured using Euclidean distance and tested with three k values:

- k = 3
- k = 5
- k = 7

Testing multiple k values aimed to identify the optimal parameter that yields the highest classification accuracy.



Pig 4. Implementation

3.5 Testing Results

The evaluation results of the K-Nearest Neighbor (K-NN) model indicate that the selection of the k value has a significant impact on classification accuracy. The experiments were conducted using three different k values, namely 3, 5, and 7. The accuracy results for each experiment are presented in the following table.

Tabel 2. Accuracy Table

k Value	Accuracy
3	0.90
5	0.93
7	0.87

Based on the table, k = 5 achieved the highest accuracy of 93% and is therefore considered the optimal parameter for this dataset. A smaller k value (e.g., k = 3) makes the model more sensitive to specific data points, while a larger k value (e.g., k = 7) reduces specificity, leading to lower accuracy.

Confusion Matrix k = 5

Table 3. Predicted

Actual \ Predicted	Apple	Orange	Grape
Apple	2	0	0
Orange	1	1	0
Grape	0	1	1

Based on the confusion matrix, the K-NN model is able to classify the apple class very well, as all apple samples in the test dataset are correctly predicted without any errors. For the orange class, two samples were tested, but only one was correctly classified, while the other was misclassified as an apple, indicating similarity in feature values between oranges and small-sized apples. For the grape class, one sample was correctly predicted, whereas the other was misclassified as an orange. This misclassification is caused by the close similarity in size and weight ranges between large grapes and small oranges. Overall, the confusion matrix results indicate that the model performs reasonably well in distinguishing each fruit class, with most misclassifications occurring among classes that have closely related physical characteristics, such as oranges and grapes.

Conclusion

Based on the results of this study, the application of the K-Nearest Neighbor (K-NN) algorithm has proven to be effective in classifying fruit types based on three main features, namely color, size, and weight, allowing the research objectives to be achieved optimally. The model, developed using 100 simulated fruit data samples, is able to identify characteristic patterns among fruit categories such as apples, oranges, and grapes, achieving a highest accuracy of 93%. This result indicates that the K-NN algorithm performs very well on small to medium-sized datasets. From a theoretical perspective, this study contributes to strengthening the use of machine learning as a classification method based on numerical and visual features. From a practical perspective, the results can serve as a foundation for the development of faster, more efficient, and more accurate fruit sorting systems in the fruit processing industry. However, this study still has limitations, particularly related to the limited dataset size and the use of relatively simple features, which means that the model performance can be further improved. Therefore, future research is recommended to increase the dataset size, incorporate image-based or texture-based features, evaluate comparative algorithms such as Support Vector Machine (SVM) or Random Forest, and integrate the system with Internet of Things (IoT) devices to support real-time automated fruit sorting.

References

[1] M. S. Irfanuddin, "Klasifikasi Tingkat Kematangan Belimbing Madu Berdasarkan Karakteristik Warna Menggunakan Algoritma K-Nearest Classification Of Maturity Level Of Honey Martumber Based On Color Characteristics Using The K-Nearest Neighbor Algorithm," vol. 4, no. 2, 2024, doi: 10.24176/detika.v4i2.12804.

[2] L. Farokhah and P. Korespondensi, "Implementasi K-Nearest Neighbor Untuk Klasifikasi Bunga Implementation Of K-Nearest Neighbor For Flower Classification With Extraction Of Rgb Color Features," vol. 7, no. 6, 2020, doi: 10.25126/jtiik.202072608.

[3] W. Saputro, D. B. Sumantri, S. Tinggi, I. Komputer, and C. Karya, "Digital Image Implementation In Grape Classification With K-," vol. 5, pp. 248–253, 2022.

[4] Y. Reswan, R. Toyib, H. Witriyono, and A. Anggraini, "Klasifikasi Tingkat Kematangan Buah Nanas Berdasarkan Fitur Warna Menggunakan Metode K – Nearest Neighbor (KNN)," vol. 20, no. 1, pp. 280–287, 2024.

[5] F. Yusuf, A. L. Fadillah, and M. R. Naufal, "K-Nearest Neighbor Untuk Klasifikasi Jenis Buah Berdasarkan Berat, Tinggi, dan Lebar," vol. 9, no. 10, pp. 703–708, 2023.

[6] V. N. Juli and P. Astuti, "Computer Science (CO-SCIENCE) Klasifikasi Kualitas Buah Apel Dengan Algoritma K-Nearest Neighbor (K-NN) Menggunakan Bahasa Pemrograman Python," vol. 4, no. 2, pp. 127–132, 2024.

[7] S. Nurjanah, A. Ramadhan, dan M. H. Kurniawan, "*Klasifikasi Jenis Buah Tomat Menggunakan Algoritma K-Nearest Neighbor dan Support Vector Machine*," pp. 800–807, 2025.

[8] R. Firmansyah dan I. P. Sari, "*Klasifikasi Tingkat Kematangan Buah Sawit Berdasarkan Ekstraksi Fitur RGB dan GLCM Menggunakan Metode K-NN*," vol. 22, pp. 457–466, 2023.

[9] S. Kardena, F. Izzati, and S. Jakarta, "Classification of Coconut Fruit Quality Using The K-Nearest Neighbour (K-NN) Method Based on Feature Extraction : Color ," vol. 18, no. 1, pp. 143–156, 2025.

[10] W. W. Widiyanto, E. Purwanto, and K. Neighbor, "Classification Of Mango Fruit Quality Based On Texture Characteristics Of Glcm (Gray Level Co-Occurrence Matrices) With Algorithm K-NN (K-NEAREST NEIGHBORS)," vol. 20, no. 1, pp. 1–10, 2019.

[11] S. Yao, "Machine Learning in Agriculture : KNN-Based Classification of Fruits and Vegetables".

[12] E. Aenun, N. Munfaati, and A. Witanti, "Klasifikasi Buah dan Sayuran Segar atau Busuk Menggunakan Convolutional Neural Network," vol. 9, no. 1, pp. 27–38, 2024.

[13] C. Aditya, N. Fadlilah, S. Surorejo, and Z. Arif, "Klasifikasi Jenis Buah Apel Menggunakan Algoritma Convolutional Neural Network," vol. 4, no. 3, pp. 5201–5208, 2025.

[14] I. H. Ikasari, P. Rosyani, and R. Amalia, "Klasifikasi Jenis Buah Menggunakan Metode CNN," vol. 4, no. 2, pp. 5451–5458, 2025.

[15] M. Muchtar *et al.*, "Perbandingan Metode K-NN dan SVM untuk Klasifikasi Kematangan Buah Mangga Berdasarkan Citra HSV dan Fitur Statistik" vol. 12, no. 2, pp. 876–884, 2024.

[16] Y. Setiawan, W. Astuti, and M. R. R. Saelan, "Classification for Papaya Fruit Maturity Level With

Convolutional Neural Network," vol. 5, no. 3, 2023.
[17] V. N. Juli, A. Ollivia, and C. Pratiwi, "Klas ifikasi Jenis Anggur Berdasarkan Bentuk Daun Menggunakan Convolutional Neural Network Dan K-Nearest Neighbor," vol. 3, no. 2, 2023.
[18] K. Kematangan, B. Apel, B. Warna, and M. Metode, "Classification of apple maturity based on color using the K-Nearest Neighboor (KNN) method," vol. 21, no. 1, pp. 55–64, 2024, doi: 10.31515/telematika.v21i1.11773.
[19] S. Informatika *et al.*, "Klasifikasi Kematangan Buah Apel dengan Ekstraksi Fitur Haralick dan KNN," vol. 1, pp. 1085–1092, 2023.
[20] M. K. Neighbors, N. Wijaya, and A. Ridwan, "Klasifikasi Jenis Buah Apel Dengan," vol. 08, pp. 74–78, 2019.
[21] V. N. Juli and P. Astuti, "Computer Science (CO-SCIENCE) Klasifikasi Kualitas Buah Apel Dengan Algoritma K-Nearest Neighbor (K-NN) Menggunakan Bahasa Pemrograman Python," vol. 4, no. 2, pp. 127–132, 2024.