Electronic Integrated Computer Algorithm Journal
Vol. 2, No. 2, April 2025, pp 91~104

ISSN: 3031-0350, DOI: 10.62123/enigma.v2i2.60

Modification of K-Nearest Neighbor Method with Normalized Euclidean Distance for Classification of Local Berastagi Orange Quality

Ananda Afifah Siregar[1], Al-Khowarizmi[1*]

[1] *Department of InformationTechnology, Universitas Muhammadiyah Sumatera Utara,* Indonesia.

DOI : 10.62123/enigma.v2i2.60

Received : March 07, 2025
Revised : March 20, 2025
Accepted : March 24, 2025

Keywords:
Classification, Berastagi Local Orange, K-Nearest Neighbor, Normalized Euclidean Distance, Matlab

ABSTRACT

Local Indonesian fruit is one example of Indonesia's natural wealth, one of which is the local Berastagi orange. Oranges are rich in vitamin C which is good for body health. Oranges tend to have a sour, fresh, and sweet taste. The vitamin C contained in oranges is 97.3 milligrams or equivalent to 163% of the nutritional adequacy rate. Not only Vitamin C, oranges also contain vitamin B6, antioxidants and fiber. Therefore, it is highly recommended to consume oranges every day because oranges can facilitate digestion, reduce the risk of diabetes, maintain healthy skin, and also maintain endurance. This study aims to apply the Classification and assessment of the quality of local oranges using the K-Nearest Neighbor (KNN) method modified with Normalized Euclidean distance to classify the quality of local Berastagi oranges based on the color of the fruit image. The research dataset was taken from 100 images of local Berastagi oranges, where the 100 images were divided into 2, namely, good oranges and bad oranges. The classification process for local Berastagi oranges uses the matlab application.

1. INTRODUCTION

Along with the development of science and information technology at this time, which is developing quite rapidly [1]. Increasingly triggered to use advanced technology as a tool or media to maintain its existence in society [2]. The impact of the increasingly rapid development of information technology, humans can determine the quality of local oranges by designing a citrus fruit quality classification system by utilizing today's rapidly developing technology. Data mining is a process that aims to find relationships in data that are unknown to users and present them in a way that is easy for users to understand so that these relationships can be the basis for decision making [3]. There are several methods or algorithms in data mining that can carry out the classification process, one of which is the K-Nearest Neighbor (KNN) algorithm [4]. K-Nearest Neighbors (KNN) is a machine learning method that uses historical data to predict future results [5]. The K-Nearest Neighbor algorithm is not only used for forecasting, but also for predicting classification [6]. KNN is the result of further improvisation of the Nearest Neighbor classification technique. This is evident because each new data can be analyzed based on its k nearest neighbors, where k is a positive integer, even with a small dataset [7]. K-Nearest Neighbor (KNN) in performing classification generally uses the Euclidean distance formula [8].

An image is a picture or similarity of an object. An image is the output of a data recording system which is usually in the form of a photo, digital in nature that can be directly stored on a storage media [9]. Images are generally used in the fruit classification process. In image classification processing, several image characteristics are generally needed. At the level of the image extraction process, it is the delivery of the image that is very important to know the classification results of the image. There are several types of extraction in image feature extraction, namely the presence of color. To determine the quality of good oranges is based on the color studied [10]. Oranges are fruits that belong to the genus Citrus. Oranges are rich in vitamin C which is good for body health. Oranges tend to have a sour, fresh, and sweet taste. Oranges are usually orange in color, but there are also yellow and green ones. The vitamin C contained in oranges is 97.3 milligrams or equivalent to 163% of the nutritional adequacy from Kompas 2023. In selecting oranges that have good quality, it can be done manually or by guessing the fruit, whether the oranges that are sorted are good for consumption. When selecting good quality oranges, manual observation is carried out by examining the color of the fruit. Selecting oranges manually provides less effective classification results due to the possibility of sorting errors from the human eye.

Based on the problems above, a solution is needed that can help in classifying the quality of local oranges. Hereby the author will provide an overview of the existence of a citrus fruit classification processing system using the K-Nearest Neighbor (KNN) method that can provide solutions and solve problems [11]. So that it makes it easier for the author to determine the quality of local citrus fruits into the 'good quality' and 'bad quality' groups, which the author describes in this paper.

2. RESEARCH METHODS

2.1 Method

*Corresponding Author Email: alkhowarizmi@umsu.ac.id

The data collection methods used in this study are observation and literature study. Observations were carried out directly at the Berastagi fruit tax. Literature study was carried out by collecting data and information from books, journals, and articles related to classification. The observation method is a data collection technique carried out through systematic observation and recording of the phenomena being investigated, observations are not limited to observations made both directly and indirectly [12].

2.2 Dataset

In this study, the author has a dataset of 100 orange images, where the 100 images are divided into 2, namely good oranges and bad oranges.

Table 1. Dataset

	Data	Picture
1	Good Orange	
..	…..	……….
..	…..	……….
100	Good Orange	

2.3 Normalized

Normalized is the process of adjusting the scale of features or input variables so that they are all in the same value range, usually between 0 and 1. The dataset used is an image of an orange. The normalization method used in this study is: Min-max normalized is a normalization method by performing a linear transformation on the original data so that it produces a balance of comparative values between data before and after the process [13]. This method can use the formula:

$$N = \frac{n - Xmin}{Xmax - Xmin} \quad (1)$$

n = original value
Xmin = lowest value
Xmax = highest value

Figure 1. Example of Orange Fruit Normalization

2.4 Research Procedures

This research has several stages, namely image acquisition, preprocessing, feature extraction, and classification. Image acquisition uses a smartphone camera and a Nikon camera. The preprocessing is to change the image size. The feature extraction used is the color of the fruit. Image classification uses the K-Nearest Neighbor algorithm [14].

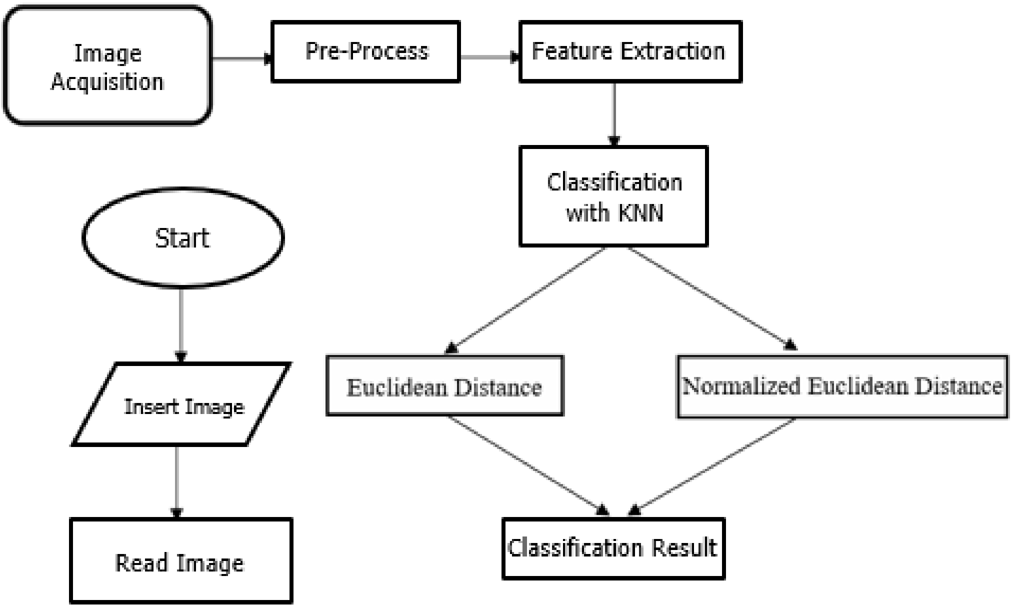


Figure 2. Research Stages Process

1. Image acquisition
   The initial stage before processing, this stage is the process of taking images using a smartphone camera and a Nikon camera.
2. Pre-process
   Pre-process is the stage of image processing to obtain good image quality. Some stages in the pre-process are resizing and cropping to reduce the size of the image.
3. Feature extraction
   At this stage, the image feature extraction used is the image color, by taking the RGB (Red, Green, and Blue) values to calculate the Euclidean distance.
4. Classification with KNN

$$D(a,b) = \sqrt{\sum_{k=1}^{d} (a_k - b_k)^2} \tag{2}$$

Classification using the KNN method with the Euclidean distance formula

$$D_{norm}(a,b) = \sqrt{\sum_{k=1}^{d} \left(\frac{a_k - b_k}{\text{range}_k} \right)^2} \tag{3}$$

D(a,b) : scalar distance of two data vectors a and b
D : dimension size
K : number of nearest neighbors
Range k : range (maximum value – minimum value) of the i-th dimension in the dataset.

5. Classification results
   From the classification results, it can be decided whether the K-Nearest Neighbor (k-NN) method in this study can classify orange images of "good" or "bad" quality.

2.5 Classification System Flowchart

In the image below, the flowchart of the classification system illustrates the system's workflow to determine the quality of oranges that are input into the system according to the orange image characteristic data contained in the system. The systematic flowchart below is, starting with inputting an orange image. After the image is input, the system will read the image. After that, the system will extract the characteristics of the image that has been input, then the system will determine the classification of the quality of the orange, if the orange has good characteristics, then the quality of the orange is good, and vice versa if the orange has bad characteristics, then the quality of the orange is not good.

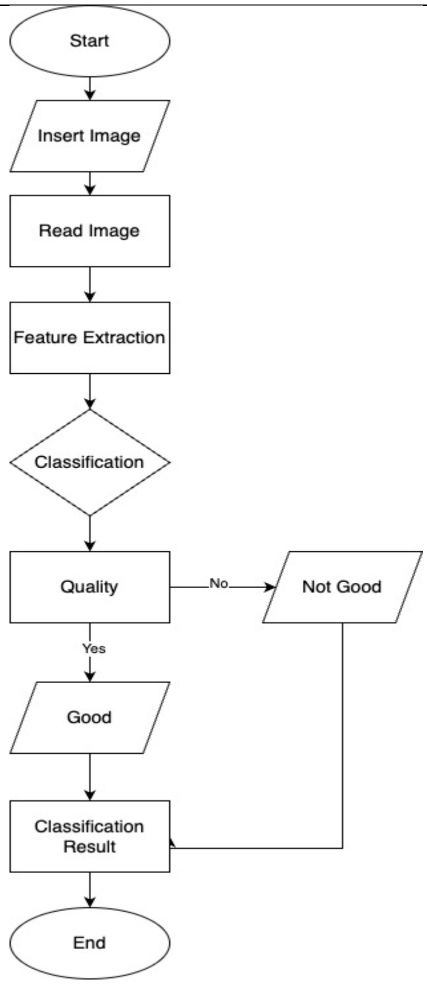


Figure 3. Classification System Flowchart

3. RESULTS AND DISCUSSIONS

In this study, the classification of the quality of local Berastagi oranges was carried out using the KNN (K-Nearest Neighbor) method with Normalized Euclidean Distance. The features used are the color features of R, G, B values. After the identification is complete, the evaluation of the results is carried out by measuring the quality level of local Berastagi oranges from the developed system.

3.1 System View

Figure 4. KNN System View

The system is made to classify the quality of local Berastagi oranges using Matlab. The system will display a column to input the image sample to be tested, then in the crop image menu the user can crop the image sample, there are 2 options, namely rectangle and square. In the image segmentation menu, the segmentation results will come out, the feature extraction menu will come out with the RGB value results. Then in the KNN classification menu, it can be done with Euclidean Distance and Normalized Euclidean Distance and the user can see the results of the orange sample classification [15].

3.2 RGB (Red, Green, Blue) values

The process of getting RGB values using the Matlab command is as follows:

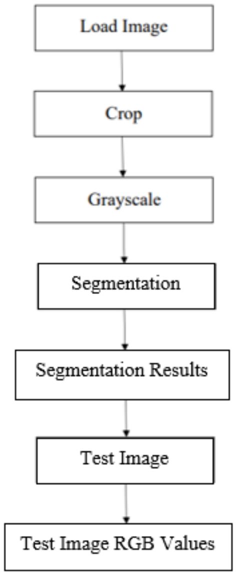


Figure 5. RGB Value Process

The user will input and crop the image to be tested, the system will read the image and then the system will change the image to grayscale. Then the system performs the segmentation process until the segmentation results. When the segmentation results are out, the system will process the segmentation results into a test image and the RGB value of the test image as shown in the image below.

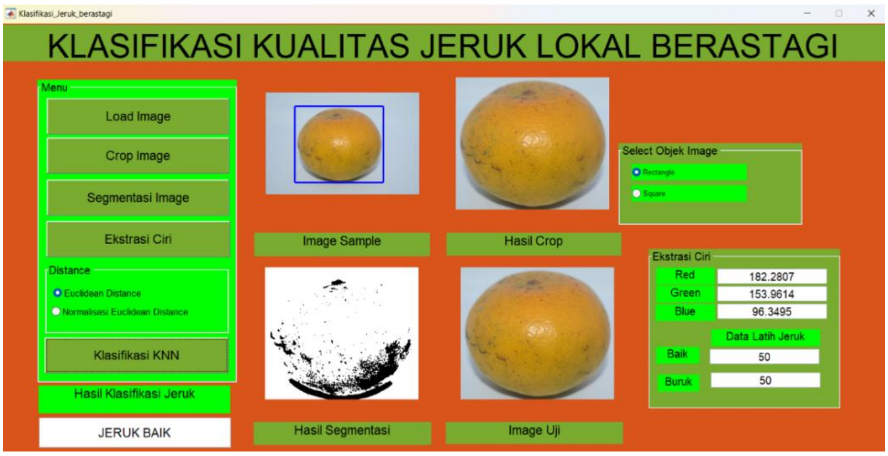


Figure 6. Sample Test Example 3

For the other 19 samples, see the table below:

Table 2. Orange Data Samples

ITEM	ORANGE DATA	RED	GREEN	BLUE	QUALITY
1	Sample 1	158.7066	131.8587	80.445	GOOD
2	Sample 2	166.4275	138.94	97.3779	BAD
3	Sample 4	164.4274	130.3507	73.7947	GOOD
4	Sample 5	155.8918	126.3104	64.4122	GOOD
5	Sample 6	185.1111	138.6269	84.5178	BAD

6	Sample 7	137.2475	104.47	64.4259	BAD
7	Sample 8	134.1979	101.6341	64.4494	BAD
8	Sample 9	173.6608	136.2476	75.5207	GOOD
9	Sample 10	122.9824	88.7374	55.6396	BAD
10	Sample 11	194.9772	146.1809	85.2512	GOOD
11	Sample 12	152.5341	121.4093	72.609	GOOD
12	Sample 13	158.397	138.8766	73.1123	BAD
13	Sample 14	165.9128	140.7492	83.4876	GOOD
14	Sample 15	161.864	142.511	86.3462	BAD
15	Sample 16	133.0953	130.2986	90.4749	BAD
16	Sample 17	177.4837	137.2362	80.6419	GOOD
17	Sample 18	173.557	137.1082	77.9342	GOOD
18	Sample 19	138.7763	127.7308	81.5836	BAD
19	Sample 20	129.1697	93.7282	48.5259	BAD

3.3 Euclidean Distance

Table 3. Samples to be Tested

ITEM	DATA ORANGE	RED	GREEN	BLUE	QUALITY
1	Sample 1	158.7066	131.8587	80.445	GOOD
2	Sample 2	166.4275	138.94	97.3779	BAD
3	Sample 3	182.2807	153.9614	96.3495	GOOD
4	Sample 4	164.4274	130.3507	73.7947	GOOD
5	Sample 5	155.8918	126.3104	64.4122	GOOD
6	Sample 6	185.1111	138.6269	84.5178	BAD
7	Sample 7	137.2475	104.47	64.4259	BAD
8	Sample 8	134.1979	101.6341	64.4494	BAD
9	Sample 9	173.6608	136.2476	75.5207	GOOD
10	Sample 10	122.9824	88.7374	55.6396	BAD
11	Sample 11	194.9772	146.1809	85.2512	GOOD
12	Sample 12	152.5341	121.4093	72.609	GOOD
13	Sample 13	158.397	138.8766	73.1123	BAD
14	Sample 14	165.9128	140.7492	83.4876	GOOD
15	Sample 15	161.864	142.511	86.3462	BAD
16	Sample 16	133.0953	130.2986	90.4749	BAD
17	Sample 17	177.4837	137.2362	80.6419	GOOD
18	Sample 18	173.557	137.1082	77.9342	GOOD
19	Sample 19	138.7763	127.7308	81.5836	BAD
20	Sample 20	129.1697	93.7282	48.5259	BAD
21	Sample 21	154.6181	148.8581	88.9745	?

Calculating the Euclidean Distance value is used to identify the oranges that are most similar to the orange being tested. KNN will calculate the distance between the unclassified orange and each orange in the classified dataset, then select the k closest oranges to determine the quality of the orange being tested based on the majority of the quality of the k nearest neighbors. The author will find the Euclidean Distance value of the orange sample with k = 3 using Equation 2.2 as follows:

$$D(a,b,c) = \sqrt{\sum_{k=1}^{d} (a_k - b_k - c_k)^2}$$

$$D(a,b,c) = \sqrt{(158.7066 - 154.6181)^2 + (131.8587 - 148.8581)^2 + (80.445 - 88.9745)^2}$$

$$D(a,b,c) = \sqrt{(4.0885)^2 + (-16.9994)^2 + (-8.5295)^2}$$

$$D(a,b,c) = \sqrt{(16.71583225) + (288.97960036) + (72.75237025)}$$

$$D(a,b,c) = \sqrt{378.44780286}$$

$$D(a,b,c) = 19.4537349334$$

Then after the Euclidean Distance value is calculated, the value will be sorted based on the smallest value. For the Euclidean Distance value of 20 other samples, it can be seen in the table below:

Table 4. Euclidean Distance

ITEM	DATA ORANGE	RED	GREEN	BLUE	EUCLIDEAN DISTANCE	SEQUENCE	CLASSIFICATION
1	Sample 1	158.7066	131.8587	80.445	19.45373493	5	GOOD
2	Sample 2	166.4275	138.94	97.3779	17.5626811	3	BAD
3	Sample 3	182.2807	153.9614	96.3495	29.08012611	12	GOOD
4	Sample 4	164.4274	130.3507	73.7947	25.86836967	7	GOOD
5	Sample 5	155.8918	126.3104	64.4122	33.36656513	15	GOOD
6	Sample 6	185.1111	138.6269	84.5178	32.47095128	14	BAD
7	Sample 7	137.2475	104.47	64.4259	53.61599508	17	BAD
8	Sample 8	134.1979	101.6341	64.4494	56.99623912	18	BAD
9	Sample 9	173.6608	136.2476	75.5207	26.50761906	8	GOOD
10	Sample 10	122.9824	88.7374	55.6396	75.67385044	20	BAD
11	Sample 11	194.9772	146.1809	85.2512	40.61880495	16	GOOD
12	Sample 12	152.5341	121.4093	72.609	32.02513494	13	GOOD
13	Sample 13	158.397	138.8766	73.1123	19.11857255	4	BAD
14	Sample 14	165.9128	140.7492	83.4876	14.94759442	2	GOOD
15	Sample 15	161.864	142.511	86.3462	9.98482379	1	BAD
16	Sample 16	133.0953	130.2986	90.4749	28.45939494	11	BAD
17	Sample 17	177.4837	137.2362	80.6419	26.96917584	9	GOOD
18	Sample 18	173.557	137.1082	77.9342	24.8722799	6	GOOD
19	Sample 19	138.7763	127.7308	81.5836	27.42172196	10	BAD
20	Sample 20	129.1697	93.7282	48.5259	72.95900341	19	BAD
21	Sample 21	154.6181	148.8581	88.9745	0		?

K = 3 BAD

3.4 Normalized Euclidean Distance

Table 5. Values to be changed

NO	DATA ORANGE	RED	GREEN	BLUE	QUALITY
1	Sample 1	158.7066	131.8587	80.445	GOOD
2	Sample 2	166.4275	138.94	97.3779	BAD
3	Sample 3	182.2807	153.9614	96.3495	GOOD
4	Sample 4	164.4274	130.3507	73.7947	GOOD
5	Sample 5	155.8918	126.3104	64.4122	GOOD
6	Sample 6	185.1111	138.6269	84.5178	BAD
7	Sample 7	137.2475	104.47	64.4259	BAD

8	Sample 8	134.1979	101.6341	64.4494	BAD
9	Sample 9	173.6608	136.2476	75.5207	GOOD
10	Sample 10	122.9824	88.7374	55.6396	BAD
11	Sample 11	194.9772	146.1809	85.2512	GOOD
12	Sample 12	152.5341	121.4093	72.609	GOOD
13	Sample 13	158.397	138.8766	73.1123	BAD
14	Sample 14	165.9128	140.7492	83.4876	GOOD
15	Sample 15	161.864	142.511	86.3462	BAD
16	Sample 16	133.0953	130.2986	90.4749	BAD
17	Sample 17	177.4837	137.2362	80.6419	GOOD
18	Sample 18	173.557	137.1082	77.9342	GOOD
19	Sample 19	138.7763	127.7308	81.5836	BAD
20	Sample 20	129.1697	93.7282	48.5259	BAD
21	Sample 21	154.6181	148.8581	88.9745	?

The purpose of Normalized is to change the value of a numeric column in a Dataset to a scale, without changing the difference in the range of values. Normalized is also one of the stages that can improve the accuracy of the data. The normalized used is Min-Max with a range of 0-1. The author will first change the Euclidean Distance value to Normalized Euclidean Distance using Equation 2.1 as follows:

$$N = \frac{n - Xmin}{Xmax - Xmin} \quad (6)$$

$$N = \frac{158.7066 - 122.9824}{194.9772 - 122.9824}$$

$$N = \frac{35.7242}{71.9948}$$

$$N = 0.496205281$$

For the Normalized Euclidean Distance values of the other 20 samples, see the table below:

Table 6. Normalized RGB Values

ITEM	DATA ORANGE	NORMALIZATION RED	NORMALIZATION GREEN	NORMALIZATION BLUE
1	Sample 1	0.496205281	0.661126273	0.65338369
2	Sample 2	0.603447749	0.769695204	1
3	Sample 3	0.823646986	1	0.978948661
4	Sample 4	0.575666576	0.638005949	0.517252108
5	Sample 5	0.457108013	0.576060959	0.325192418
6	Sample 6	0.862960936	0.764894824	0.736753869
7	Sample 7	0.198140699	0.241208758	0.325472857
8	Sample 8	0.155782084	0.197729363	0.325953902
9	Sample 9	0.703917505	0.728415921	0.552583313
10	Sample 10	0	0	0.145617375
11	Sample 11	1	0.880711088	0.75176656
12	Sample 12	0.410469923	0.500918374	0.49298084
13	Sample 13	0.491904971	0.768723169	0.503283387
14	Sample 14	0.596298622	0.79743346	0.715665684
15	Sample 15	0.540061227	0.82444499	0.7741812
16	Sample 16	0.140467089	0.637207163	0.858695652
17	Sample 17	0.757017173	0.743572918	0.657414231
18	Sample 18	0.702475734	0.74161045	0.601987636
19	Sample 19	0.219375566	0.597838219	0.676690821
20	Sample 20	0.085940929	0.076517846	0
21	Sample 21	0.439416458	0.921757329	0.827982478

The author will search for the Normalized Euclidean Distance value of the orange sample with k = 3 using Equation 3.2 as follows:

$$D(a,b,c) = \sqrt{\sum_{k=1}^{d} (a_k - b_k - c_k)^2}$$

$$D(a,b,c) = \sqrt{(0.496205281 - 0.439416458)^2 + (0.661126273 - 0.921757329)^2 + (0.65338369 - 0.827982478)^2}$$

$$D(a,b,c) = \sqrt{(0.056788823)^2 + (-0.260631056)^2 + (-0.174598788)^2}$$

$$D(a,b,c) = \sqrt{(0.00322497042) + (0.06792854735) + (0.03048473677)}$$

$$D(a,b,c) = \sqrt{0.10163825454}$$

$$D(a,b,c) = 0.318807551 \quad (7)$$

Then after the Normalized Euclidean Distance value is calculated, the value will be sorted based on the smallest value. For the Normalized Euclidean Distance value of the other 20 samples, it can be seen in the table below:

Table 7. Normalized Euclidean Distance

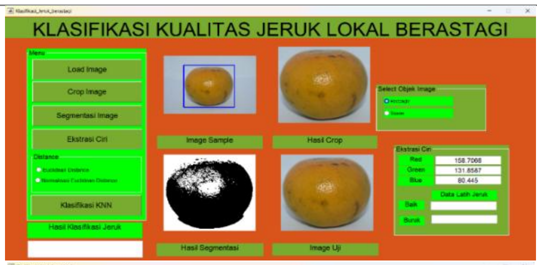
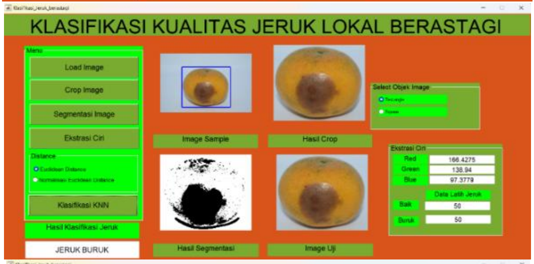
NORMALIZATION EUCLIDEAN	SEQUENCE	CLASSIFICATION
0.318807551	4	GOOD
0.282168712	3	BAD
0.420173539	10	GOOD
0.442303406	12	GOOD
0.610423472	16	GOOD
0.46078024	13	BAD
0.879702375	17	BAD
0.925579626	18	BAD
0.428002808	11	GOOD
1.228147194	19	BAD
0.567227968	15	GOOD
0.538673765	14	GOOD
0.362772653	5	BAD
0.229528865	2	GOOD
0.149978793	1	BAD
0.413863288	8	BAD
0.402136078	7	GOOD
0.390802699	6	GOOD
0.419798462	9	BAD
1.234880446	20	BAD
0		

K = 3 BAD

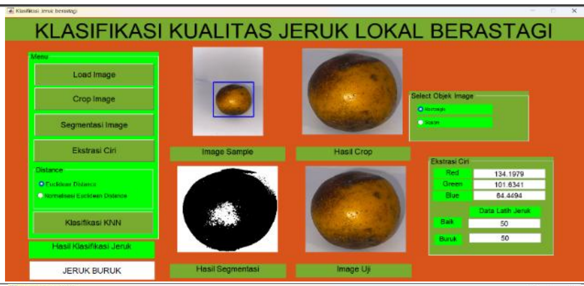
3.5 Test Results

In this study, the author conducted a test of 20 local Berastagi orange image data. The results of the data testing can be seen in the table below:

Table 8. Test Results

DATA ORANGE	RESULT
Sample 1	
Sample 2	
Sample 3	
Sample 4	
Sample 5	
Sample 6	
Sample 7	

Sample 8



Sample 9

Sample 10

Sample 11

Sample 12

Sample 13

Sample 14

Sample 15

Sample 16

Sample 17

Sample 18

Sample 19

Sample 20

From the 20 test samples above, the author obtained the following test results:

1) In sample 1, it has a red value of 158.7066, green 131.8587, blue 80.445 with good orange classification results.
2) In sample 2, it has a red value of 166.4275, green 138.94, blue 97.3779 with bad orange classification results.
3) In sample 3, it has a red value of 182.2807, green 153.9614, blue 96.3495 with good orange classification results.
4) In sample 4, it has a red value of 164.4274, green 130.3507, blue 73.7947 with good orange classification results.
5) In sample 5, it has a red value of 155.8918, green 120.3104, blue 64.4122 with good orange classification results.
6) In sample 6, it has a red value of 185.1111, green 138.6269, blue 84.5178 with bad orange classification results.
7) In sample 7, it has a red value of 137.2475, green 104.47, blue 64.4259 with bad orange classification results.
8) In sample 8, it has a red value of 134.1979, green 101.6341, blue 64.4494 with bad orange classification results.
9) In sample 9, it has a red value of 173.6608, green 135.2476, blue 75.5207 with good orange classification results.
10) In sample 10, it has a red value of 122.9824, green 88.7374, blue 55.6396 with poor orange classification results.
11) In sample 11, it has a red value of 194.9772, green 146.1809, blue 85.2512 with good orange classification results.
12) In sample 12, it has a red value of 152.5341, green 121.4093, blue 72.609 with good orange classification results.
13) In sample 13, it has a red value of 158.397, green 138.8766, blue 73.1123 with poor orange classification results.
14) In sample 14, it has a red value of 165.9128, green 140.7492, blue 83.4876 with good orange classification results.
15) In sample 15, it has a red value of 161.864, green 142.511, blue 86.3462 with bad orange classification results.
16) In sample 16, it has a red value of 133.0953, green 130.2986, blue 90.4749 with bad orange classification results.
17) In sample 17, it has a red value of 177.4837, green 137.2362, blue 80.6419 with good orange classification results.
18) In sample 18, it has a red value of 173.557, green 137.1082, blue 77.9342 with good orange classification results.
19) In sample 19, it has a red value of 138.7763, green 127.7308, blue 81.5836 with bad orange classification results.
20) In sample 20, it has a red value of 129.1697, green 93.7282, blue 48.5259 with bad orange classification results.
21) In the test results that the researcher conducted above, the researcher obtained an accuracy level of 88% at a neighborhood value of K = 3.

3. CONCLUSION

Based on the results and discussion in this study, it can be concluded that classifying the quality of local Berastagi oranges based on the color of the oranges using K-Nearest Neighbor starts from calculating the Euclidean Distance and Normalized Euclidean Distance to determine the distance value between 2 points. The results of the classification of local Berastagi oranges are based on finding the label of the K nearest neighbors. Based on the tests carried out, the level of accuracy of the Euclidean Distance decreased after normalization, because normalization only guarantees that all values will have the same scale, namely 0-1 but does not handle outliers well, thus affecting the performance of K-Nearest Neighbor in attracting irrelevant nearest neighbors, causing errors in the classification process. The K-Nearest Neighbor classification method can classify the quality of local Berastagi oranges with an accuracy level of 88% at a neighborhood value of K = 3.

REFERENCES

[1] E. Fauzi, M. V. Sinatrya, N. D. Ramdhani, R. Ramadhan, and Z. M. R. Safari, "Pengaruh kemajuan teknologi informasi terhadap perkembangan akuntansi," *Jurnal riset pendidikan ekonomi*, vol. 7, no. 2, pp. 189–197, 2022.

[2] A. R. P. Misar, "Penerapan Dan Penggunaan Teknologi Maju Bagi Industri Kreatif Desain Grafis Dalam Menghadapi Asean Economic Community 2015".

[3] B. H. Pangestu, "Data Mining Menggunakan Algoritma Naive Bayes Classifier Untuk Evaluasi Kinerja Karyawan," *Jurnal Riset Matematika*, pp. 177–184, 2023.

[4] I. P. N. Fajarini, I. M. I. Subroto, and A. Riansyah, "Klasifikasi Kepakaran Reviewer Menggunakan Algoritma K-Nearest Neighbor (KNN)," in *Prosiding Seminar Riset Mahasiswa*, 2022, pp. 272–280.

[5] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *J Big Data*, vol. 11, no. 1, p. 113, 2024.

[6] S. Zhang, "Challenges in KNN classification," *IEEE Trans Knowl Data Eng*, vol. 34, no. 10, pp. 4663–4675, 2021.

[7] R. S. Al-Khowarizmi, M. K. M. Nasution, and M. Elveny, "Sensitivity of MAPE using detection rate for big data forecasting crude palm oil on k-nearest neighbor," *Int. J. Electr. Comput. Eng*, vol. 11, no. 3, pp. 2696–2703, 2021.

[8] B. V. Jayadi, T. Handhayani, and M. D. Lauro, “Perbandingan Knn Dan Svm Untuk Klasifikasi Kualitas Udara Di Jakarta,” *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 11, no. 2, pp. 1–7, 2023.

[9] H. Sunandar, “Perbaikan kualitas Citra Menggunakan Metode Gaussian Filter,” *MEANS (Media Informasi Analisa dan Sistem)*, pp. 19–22, Jun. 2017, doi: 10.54367/means.v2i1.18.

[10] S. Maulana, “Klasifikasi Kualitas Buah Pir Berdasarkan Warna Dan Tekstur Dengan Menggunakan Metode K-Nearest Neighbor (KNN),” Universitas Islam Lamongan, 2021.

[11] A. Nosseir and S. E. A. Ahmed, “Automatic Classification for Fruits’ Types and Identification of Rotten Ones using k-NN and SVM.,” *International Journal of Online & Biomedical Engineering*, vol. 15, no. 3, 2019.

[12] N. F. Munazhif, G. J. Yanris, and M. N. S. Hasibuan, “Implementation of the K-nearest neighbor (kNN) method to determine outstanding student classes,” *Sinkron: jurnal dan penelitian teknik informatika*, vol. 7, no. 2, pp. 719–732, 2023.

[13] A. R. Lubis and M. Lubis, “Optimization of distance formula in K-Nearest Neighbor method,” *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 1, pp. 326–338, 2020.

[14] Y. Dang, N. Jiang, H. Hu, Z. Ji, and W. Zhang, “Image classification based on quantum K-Nearest-Neighbor algorithm,” *Quantum Inf Process*, vol. 17, pp. 1–18, 2018.

[15] H. A. Abu Alfeilat *et al.*, “Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review,” *Big Data*, vol. 7, no. 4, pp. 221–248, Dec. 2019, doi: 10.1089/big.2018.0175.