E-ISSN: 2962-6110
P-ISSN: 2964-7843
JSRET
jsret.knpub.com
Journal of Scientific Research, Education, and Technology

Implementation of The K-Means Clustering Algorithm in Determining the Rate of Indramayu Mango Fruit

Vito Dwi Yanto, Irma Handayani
Universitas Teknologi Yogyakarta, Sleman, Indonesia

ABSTRACT

This research aims to classify the ripeness levels of Indramayu mangoes using the K-Means Clustering algorithm based on HSV (Hue, Saturation, Value) color features. The process begins with capturing mango images, followed by preprocessing steps such as normalization and resizing to enhance image quality. Next, color feature extraction is conducted, focusing on the Hue value as an indicator of color changes that characterize ripeness levels. The optimal number of clusters is determined using the Elbow method, resulting in two clusters: ripe mangoes and unripe mangoes. The clustering quality evaluation is performed using the Silhouette Score, which indicates an accuracy of 80%. The results demonstrate that the K-Means algorithm successfully classifies Indramayu mangoes, generating 495 image data divided into two main categories. This study contributes to improving the efficiency of automated mango ripeness classification, with potential applications in the agricultural industry.

Keywords: *K-Means Clustering, Color Feature Extraction, Silhouette Score.*

Corresponding author
Name: Vito Dwi Yanto
Email: vitody03@gmail.com

INTRODUCTION

Indonesians are familiar with mangoes as seasonal fruits, making them widely recognized, including the Indramayu mango. Indramayu mangoes have a similar appearance when ripe and unripe. This mango variety has slightly coarse fruit fibers, contains less water compared to other types of mangoes, and offers a sweet, rich, and dry taste (Kiay, 2018). In the distribution and marketing process, issues often arise in determining the ripeness level of the fruit, particularly in distinguishing between unripe and ripe mangoes.

The identification of mango ripeness levels is still conducted manually, such as by pressing or smelling the fruit. This process is time-consuming and can reduce the fruit's quality (Premana & Pandhu Wijaya, 2022). Currently, people sort mangoes one by one, looking for differences in color and aroma. Moreover, if harvested mangoes are

already ripe and not properly sorted, they will quickly rot, making them unfit for consumption.

The K-Means Clustering algorithm is used to classify the ripeness levels of Indramayu mangoes by dividing them into two clusters: ripe and unripe. K-Means is one of the simplest clustering algorithms compared to other clustering methods (Premana & Pandhu Wijaya, 2022). This algorithm has several advantages, such as being easy to implement and execute, relatively fast, adaptable, and widely practiced in data mining tasks. Clustering is a method for grouping documents, where documents are grouped based on their content to reduce the search space needed to respond to a query (Grossman, David, and Ophir Frider, 2004). The objective of this study is to classify Indramayu mangoes by clustering mango images based on their ripeness levels using the K-Means Clustering algorithm with color features (Hue, Saturation, and Value).

Previous studies have utilized the K-Means Clustering algorithm for grouping images or visual data due to its effectiveness in unsupervised learning. Additionally, color features are often used in the clustering process for fruit ripeness levels, as changes in the fruit's surface color are a key indicator of the ripening process (Himmah, Widyaningsih, & Maysaroh, 2020). However, previous research tends to overlook image preprocessing steps such as normalization and resizing, which can lead to variations in image size and quality that affect clustering results. Moreover, prior studies have rarely employed an approach based on HSV (Hue, Saturation, Value) color feature extraction. This study addresses these issues by implementing a more systematic image preprocessing process and optimizing HSV color feature extraction.

This study employs a K-Means-based clustering method with an HSV (Hue, Saturation, Value) color feature extraction approach (Utami & Putra, 2023) to classify Indramayu mangoes into two categories: ripe and unripe. The initial stage involves image preprocessing, such as normalization and resizing, to standardize the image size and improve its quality (Premana, Saeful Bachri, & Wijaya, 2022). Subsequently, color feature extraction is performed, focusing on the average Hue value, which plays a crucial role in distinguishing the dominant color of each image. The K-Means Clustering method is then applied to divide the data into two groups: ripe mangoes and unripe mangoes. Thus, this study optimizes each process step to achieve better result.

METHOD

The research design employs a quantitative method with a descriptive approach. This approach aims to describe and interpret the research object as it is, based on actual conditions in the field (Oseko, 2017). The study is conducted to classify mango ripeness levels based on dominant color features using the K-Means Clustering algorithm. The data used consists of images of Indramayu mangoes collected from various fruit outlets.

The data for Indramayu mangoes was collected from several fruit outlets, with the research samples consisting of mango images obtained through direct

photography. The total number of images used represents the data to be processed during the preprocessing stage (Arif Rachman, 2016). The clustering results will divide the sample data into two main groups: ripe and unripe. The samples consist of 640 mango images that have undergone a cleaning process (background removal). This step ensures that the color features in the images are derived solely from the mangoes themselves and are not influenced by the background.

The K-Means algorithm is utilized for its capability in unsupervised learning to group unlabeled data by clustering similar objects into a single cluster (Febrianti, Arini, & Fahrianto, 2020). This clustering is performed based on the dominant color features extracted from mango images using the HSV (Hue, Saturation, Value) feature extraction model. The data is obtained by photographing each Indramayu mango from several fruit outlets, with lighting conditions affecting the quality of the captured images. The collected data is processed by removing the background from each image to create the dataset, which is then compiled into a complete dataset. The system's workflow can be visualized in the flowchart in Figure 1.

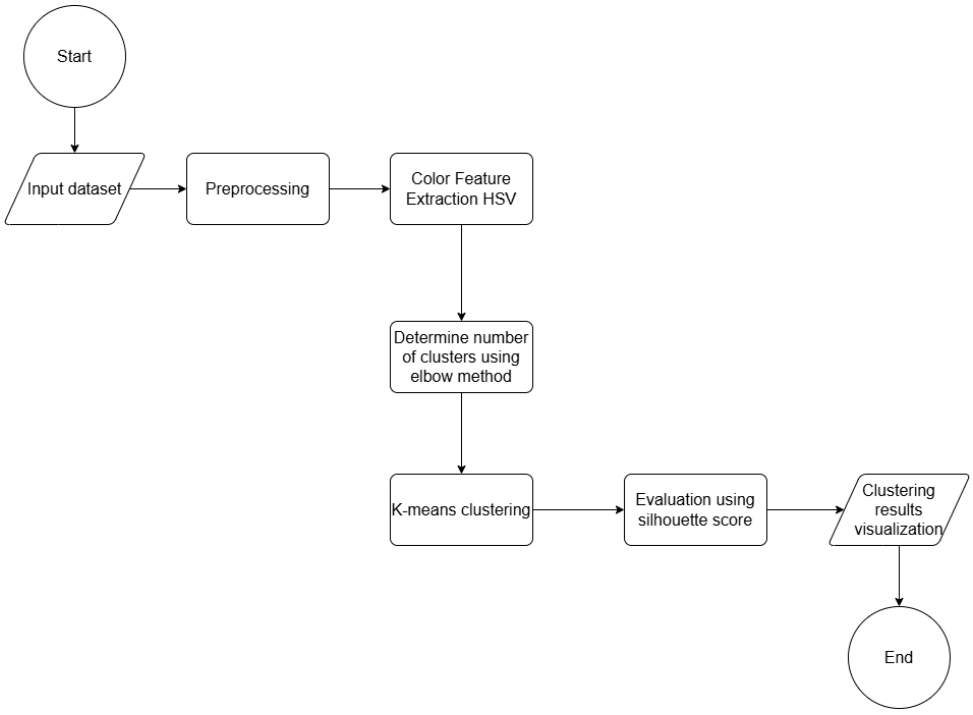


Figure 1. Flowchart diagram of the k-means clustering system

Figure 1 illustrates the workflow of the K-Means Clustering system for determining mango ripeness levels. The initial stage involves inputting a dataset of Indramayu mango images into the system, followed by data processing in the preprocessing stage. This stage aims to enhance the quality of the image data to ensure consistency and allow for more optimal analysis. The preprocessing includes resizing to

standardize image dimensions and normalization to adjust pixel intensity, reducing variations caused by lighting conditions.

After preprocessing, the next step is color feature extraction using the HSV (Hue, Saturation, Value) color space. The main focus is on the Hue value, as color changes often serve as a key indicator of mango ripeness (Muchtar & Muchtar, 2024). Subsequently, the elbow method is applied to determine the optimal number of clusters and the value of k by calculating the variance of squared differences. As k increases, the point where the rate of distortion improvement significantly decreases corresponds to the optimal k value, referred to as the "elbow" (Cui, 2016).

The process then proceeds to the K-Means Clustering stage, after determining the number of clusters using the elbow method. K-Means is applied to divide the dataset into the specified number of clusters. This algorithm works by minimizing the Euclidean distance between each data point and the cluster center (centroid) (Sinaga & Marpaung, 2020) (Sinaga & Marpaung, 2020). The results of the clustering stage are then evaluated using the Silhouette Score, which measures how well the data points are clustered. A higher score indicates better clustering performance. The clustering results are visualized in the form of 2D and 3D space graphs, providing insights into the clustering patterns of Indramayu mangoes.

FINDING AND DISCUSSION

RESEARCH RESULT

Several results were obtained from each stage of this research. First, the color feature extraction process using HSV was conducted. This process aims to extract color information from mango images, focusing on the average values of the color components in the HSV (Hue, Saturation, Value) color space (Puspa & Widayati, 2019). The RGB color space is converted into HSV, followed by calculating the average value of each HSV component for each image.

From this process, HSV results are obtained which will be displayed in the form of a table, by only taking a sample of 5 data that has been done.

Table 1: HSV result samples

NO	Hue	Saturation	Value
1	26.858889	0.150969	0.231955
2	30.084278	0.289551	0.292960
3	32.883030	0.267546	0.283439
4	22.082108	0.123897	0.215712
5	20.121723	0.137457	0.180754

The next step is data filtering by removing images with hue values that are too close to the median value. Images with such values are considered irrelevant for this

study. This filtering process distinguishes this research, as it reduces the number of images from 640 to 495. The subsequent results are illustrated in Figure 2.

```
The amount of data before cleaning: 640
The amount of data after cleaning: 495
```

Figure 2. Result of data filtering with HSV

The second process involves determining the results using the Elbow Method, with the primary goal of identifying the optimal number of clusters. The Elbow Method is a technique used to determine the optimal number of clusters (k) in the K-Means Clustering algorithm. This method plots the distortion or inertia values against the number of clusters (k). Distortion refers to the sum of squared distances between data points and their nearest centroid, also known as the within-cluster sum of squares (WCSS) (Setiady, 2020).

The elbow plot indicates the optimal number of clusters, with a significant drop observed in the elbow graph before the curve levels off (Umargono, Suseno, & Gunawan, 2020). In this study, the elbow point is identified at k = 2, resulting in two clusters: ripe mangoes and unripe mangoes. Based on the graph shown below, a significant drop is evident, as illustrated in Figure 3.

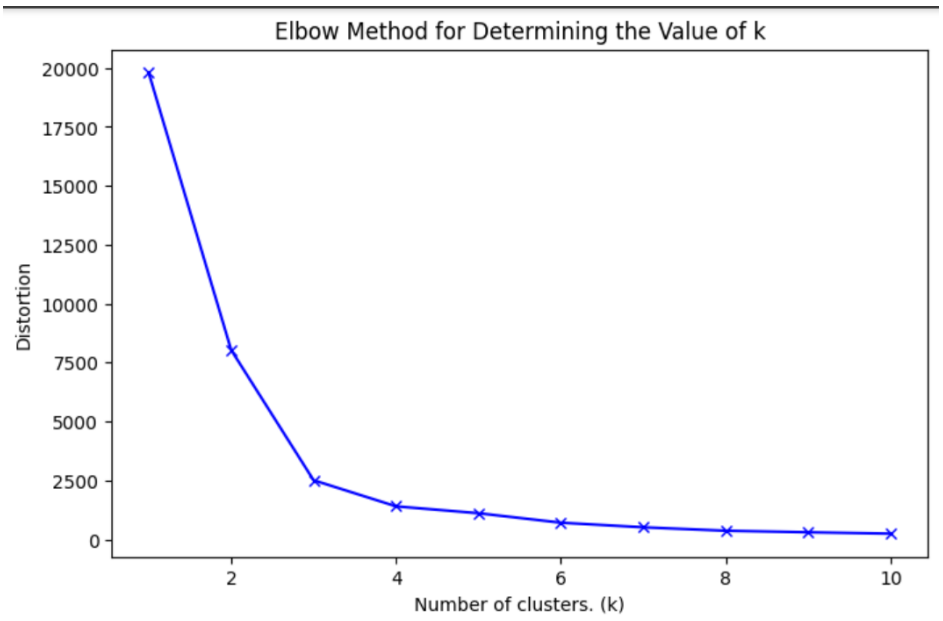
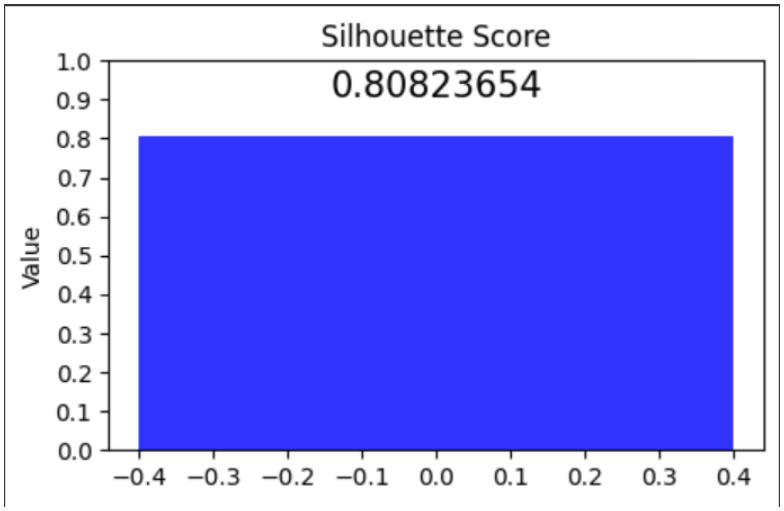


Figure 3. Results of elbow method to get k value

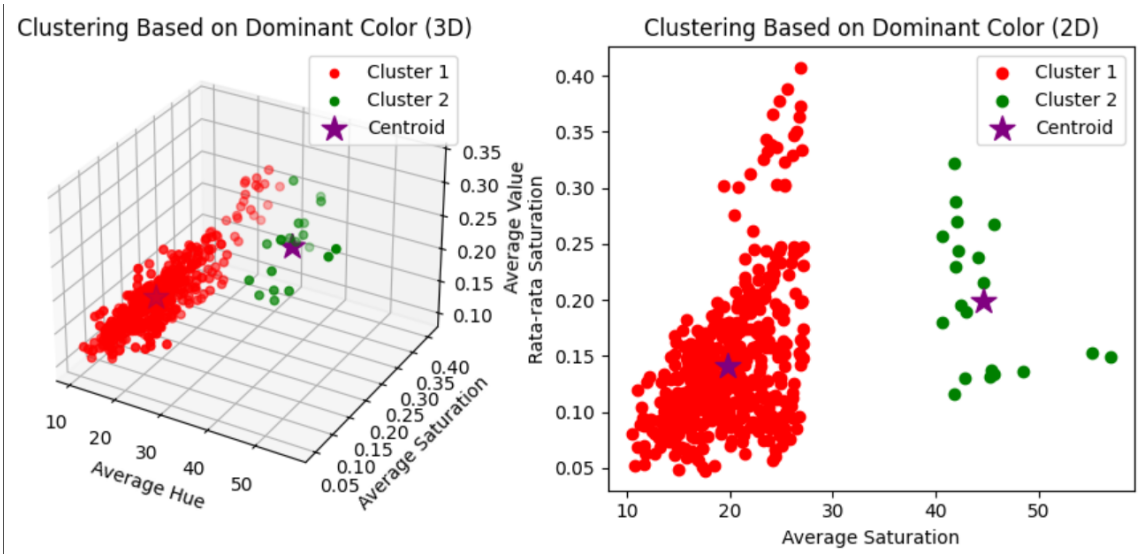
The final stage evaluates the quality of clustering using the Silhouette Score. For k = 2, the Silhouette Score indicates that the cluster division is sufficiently good. The

Silhouette Score is a reliable metric for assessing clustering quality based on cohesion and separation between clusters (Setiady, 2020).



Gambar 4. Hasil akurasi silhouette score

From the test results that have been carried out, the accuracy value of the Silhouette Score is 0.808 or 80%. This shows that K-means Clustering produces quite good results in determining the maturity level of mango fruit. From these results, the cluster results are also obtained in the form of 3D and 2D cluster position plotting, as shown in Figure 5.



Gambar 5. Hasil plotting Clustering

As shown in Figure 5, the clusters were successfully grouped into two distinct clusters: Cluster 1 represents ripe mangoes, and Cluster 2 represents unripe mangoes.

Examples of mango images successfully clustered using the K-Means algorithm can be seen in Figure 6.

Figure 6. Images successfully clustered.

Based on these results, out of the 640 data points obtained, a filtering process was conducted, leaving 495 data points. The K-Means Clustering model was successfully implemented, as evidenced by Table 2. The clustering results divided the 495 image data points into two groups:

Table 3. Indramayu mango image clustering results

NO	Status	Total
Cluster 1	Ripe	475
Cluster 2	Unripe	20

The data obtained from the fruit outlets showed that many of the Indramayu mangoes were already ripe. However, the implementation of K-Means Clustering was successfully carried out, effectively grouping the unripe and ripe mangoes.

DISCUSSION

The results of the study indicate that the K-Means Clustering algorithm is capable of grouping the ripeness levels of Indramayu mangoes with good accuracy, achieving a Silhouette Score of 0.808 or 80%. This finding is significant as it demonstrates that the HSV (Hue, Saturation, Value) color feature extraction method can be an effective solution for distinguishing ripe and unripe mangoes. The data filtering process, which removes images with hue values too close to each other, enhances the quality of the data used, resulting in more accurate clustering (Pah, Mola, & Mauko, 2021). Furthermore, preprocessing steps such as normalization and resizing

ensure data consistency, reducing variations caused by lighting or image size (Alan Pambudi, Agus Pranoto, & Panji Sasmito, 2021). The clustering results, visualized in 2D and 3D graphs, show that the two clusters, ripe and unripe mangoes, are well-separated.

The findings of this study align with previous research that highlights the effectiveness of the K-Means algorithm in clustering visual data based on color features. For example, (Himmah et al., 2020) emphasize the importance of color features in clustering fruit ripeness levels, especially since color changes on the fruit's surface serve as the primary indicator of the ripening process. However, this study addresses the weaknesses of earlier studies that tended to overlook preprocessing steps such as normalization and resizing (Premana & Pandhu Wijaya, 2022). Additionally, the HSV-based approach used in this research is more specific compared to previous studies that only utilized RGB color features (Utami & Putra, 2023). Thus, this study contributes further to the literature by refining the preprocessing steps and adopting a feature extraction approach based on HSV.

Although this study yields satisfactory results, there are several limitations that need to be considered. First, the data used were obtained from only a few specific fruit outlets, so variations in lighting conditions and image quality may not fully represent the entire population of Indramayu mangoes. Second, this study only considers color features as indicators of ripeness, while other factors, such as texture or aroma, which are also relevant in determining ripeness, were not included. Third, this method relies on the quality of the images taken, so issues such as noise or uneven lighting may affect the clustering results.

The results of this study open up opportunities for further development in the field of fruit ripeness classification. Future research could consider integrating other features, such as texture and aroma, to enhance the accuracy of clustering. Additionally, the application of other algorithms, such as DBSCAN or Gaussian Mixture Models, could be compared to evaluate the effectiveness of different clustering methods. In a practical context, the findings of this study could be applied to automate the fruit sorting process in the agricultural industry, reducing reliance on less efficient manual methods (Febrianti et al., 2020). The use of technologies such as high-resolution cameras and controlled lighting could also improve data quality, resulting in a more reliable and consistent system.

CONCLUSION

Based on the results of the study, the K-Means Clustering algorithm was able to group the ripeness levels of Indramayu mangoes with a good level of accuracy, achieving a Silhouette Score of 80%. The color feature extraction process using the HSV (Hue, Saturation, Value) color space, with a focus on the Hue value, successfully identified color changes as the main indicator of mango ripeness. This study also emphasizes the importance of preprocessing steps, such as normalization and resizing, which effectively improved image quality consistency and reduced the impact of

lighting variations. The clustering results showed that 495 image data were divided into two main groups: the ripe mango cluster (475 data) and the unripe mango cluster (20 data). The visualization of the results in 2D and 3D graphs supports the successful separation of these two groups. This study also makes an additional contribution by refining the HSV-based approach, compared to previous research that more commonly used RGB.

However, this study has several limitations. The data used is limited to a few fruit outlets, and therefore may not fully represent the entire population of Indramayu mangoes. Additionally, the study only used color features as indicators, without considering other factors such as texture or aroma. Disturbances such as noise or uneven lighting may also affect the clustering results. Future research could integrate additional features, such as texture or aroma, to improve the accuracy of the grouping. Moreover, the application of other algorithms, such as DBSCAN or Gaussian Mixture Model, could be compared to evaluate the effectiveness of different methods. Practically, the findings of this study could be implemented in an automated fruit sorting system in the agricultural industry to enhance process efficiency and reduce dependence on manual methods. The use of technologies such as high-resolution cameras and controlled lighting is also recommended to improve data quality.

REFERENCES

Alan Pambudi, N., Agus Pranoto, Y., & Panji Sasmito, A. (2021). PENGENALAN TINGKAT KEMATANGAN BUAH KOPI BERDASARKAN FITUR WARNA CIELAB DENGAN K-MEANS CLUSTERING. In *Jurnal Mahasiswa Teknik Informatika)* (Vol. 5).

Arif Rachman, MTr. H. CE. Y. A. I. S. H. P. (2016). *Kuantitatif, P. P. (2016). Metode Penelitian Kunatitatif Kualitatif dan R&D. Alfabeta, Bandung.*

Cui, M. (2016). *Introduction to the k-means clustering algorithm based on the elbow method. Accounting, Auditing and Finance, 1(1), 5-8.* https://doi.org/10.23977/accaf.2020.010102

Febrianti, A., Arini, S. T., & Fahrianto, F. (2020). PENERAPAN METODE K-MEANS CLUSTERING DAN SUPPORT VECTOR MACHINE (SVM) DALAM IDENTIFIKASI API PADA CITRA WARNA DIGITAL. *Jurnal CoreIT*, *6*(1).

Himmah, E. F., Widyaningsih, M., & Maysaroh, M. (2020). Identifikasi Kematangan Buah Kelapa Sawit Berdasarkan Warna RGB Dan HSV Menggunakan Metode K-Means Clustering. *Jurnal Sains Dan Informatika*, *6*(2), 193–202. https://doi.org/10.34128/jsi.v6i2.242

Kiay, G. S. (2018). *Konsentrasi Asam Sitrat terhadap Mutu Sari Buah Mangga Indramayu*.

Muchtar, M., & Muchtar, R. A. (2024). PERBANDINGAN METODE KNN DAN SVM DALAM KLASIFIKASI KEMATANGAN BUAH MANGGA BERDASARKAN CITRA HSV DAN

FITUR STATISTIK. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(2). https://doi.org/10.23960/jitet.v12i2.4010

Oseko, M. (2017). Knowledge Sharing and Creating Process in a Social Development Project: Developing a Model for Intra-Project Knowledge Management. *Open Journal of Social Sciences*, *05*(10), 117–127. https://doi.org/10.4236/jss.2017.510011

Pah, N. E. R., Mola, S. A. S., & Mauko, A. Y. (2021). EKSTRASI CIRI WARNA HSV DAN CIRI BENTUK MOMENT INVARIANT UNTUK KLASIFIKASI BUAH APEL MERAH. *Jurnal Komputer Dan Informatika*, *9*(2), 142–153. https://doi.org/10.35508/jicon.v9i2.5043

Premana, A., & Pandhu Wijaya, A. (2022). *Klasifikasi Jenis Buah Mangga Menggunakan Metode K-Means Clustering*. *5*.

Premana, A., Saeful Bachri, O., & Wijaya, A. P. (2022). *KLASIFIKASI JENIS MANGGA APEL MENGGUNAKAN METODE K-MEANS KLUSTERING*. Retrieved from https://jti.rivierapublishing.id/index.php/rp

Puspa, I., & Widayati, S. (2019). Segmentasi Warna Citra HSV dan Deteksi Objek Kupu-Kupu Dengan Metode Klasifikasi K-Means. *Seminar Nasional Teknologi Informasi Dan Komunikasi STI&K (SeNTIK)*, *3*(1).

Setiady, D. A. (2020). *IMPLEMENTATION OF K-MEANS ALGORITHM ELBOW METHOD AND SILHOUETTE COEFFICIENT FOR RAINFALL CLASSIFICATION* (Vol. 4).

Sinaga, A. S. R., & Marpaung, E. (2020). Segmentasi Warna HSV Telapak Tangan Untuk Deteksi Bakteri Pada Pendemi Covid 19. *Fountain of Informatics Journal*, *5*(3), 1. https://doi.org/10.21111/fij.v5i3.4925

Umargono, E., Suseno, J. E., & Gunawan, V. (2020). *K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula*.

Utami, M., & Putra, E. D. (2023). *Deteksi Objek Kualitas Daun Sawi Menggunakan Metode HSV Color dan Color Blob*. *5*(2). Retrieved from https://jurnal.ikhafi.or.id/index.php/jusibi