ATCSJ
Applied Technology and Computing Science Journal
Open Access
p-ISSN 2621-4458 | e-ISSN 2621-4474
UNUSA
Available online to journal2.unusa.ac.id
UNUSA
S4-Accredited – SK No.200/M/KPT/2021
Journal Page is available to https://journal2.unusa.ac.id/index.php/ATCSJ

# Red Dragon Fruit (*Hylocereus costaricensis*) Ripeness Color Classification by Naïve Bayes Algorithm

**Zilvanhisna Emka Fitri[1], Mega Silvia[2], Abdul Madjid[3], AMN Imron[4], Lalitya Nindita Sahenda[5]**

[1,2,3,4,5] *Politeknik Negeri Jember, Indonesia*

[1*] zilvanhisnaef@polije.ac.id

*Article history:*

Received 20 april 2022
Revised 10 May 2022
Accepted 28 May 2022
Available online 10 June 2022

*Keywords:*
Computer Vision
Dragon Fruit
GLCM
Naïve Bayes
Maturity

***Abstract***

*Dragon fruit is a unique fruit that is popular in Indonesia. besides having a sweet taste, this fruit also contains fiber, vitamins and minerals that are good for health. Dinas Pertanian Kabupaten Banyuwangi noted that the total dragon fruit production was 906,511.61 tons and the total productivity was 261.14 Kw/Ha in 2018. This shows that Kabupaten Banyuwangi is one of the largest producers of red dragon fruit in East Java Province. One of the issues determining the quality of dragon fruit is choosing the time to harvest, because dragon fruit is a climate-averse fruit. Non-climateric fruit is when we harvest fruit in its raw state, the fruit will never become ripe, so determining the harvest time for dragon fruit is very important. The determination made by paying discoloration and sizes of dragon fruit that is considered less effective. To overcome this, a system has been created that can automatically determine the ripeness of dragon fruit using digital image processing techniques and intelligent systems. The parameters used are color features and GLCM texture features using angles 0°, 45°, 90° and 135° These features are parameters in the classification process using the Naïve Bayes method. Naïve bayes can classify the ripeness of red dragon fruit (Hylocereus costaricensis) with 87.37% accuracy.*



## I. INTRODUCTION -11 PT

Dragon fruit has another name, namely Pitaya, which is currently cultivated in Southeast Asian countries such as Indonesia. The demand for dragon fruit tends to increase every year with annual productivity of 24-30 tons/ha[1]. Dinas Pertanian Kabupaten Banyuwangi noted that the total dragon fruit production was 906,511.61 tons and the total productivity was 261.14 Kw/Ha in 2018, while the dragon fruit centers are in sub-districts such as Bangorejo, Purworejo, Pesanggaran, Siliragung, Muncar and Tegal[2]. This shows that Kabupaten Banyuwangi is one of the largest producers of red dragon fruit in East Java Province. Dragon fruit is favored by the public because of its unique appearance, sweet taste, and contains fiber, vitamins, and minerals that are good for health. However, there are some issues to judge the quality of dragon fruit, such as: fruit diseases, harvest time selection, post-harvest sorting and grading processes [3].

[1] Corresponding author



DOI: https://doi.org/10.33086/atcsj.X.

Dragon fruit is also a non-climacteric fruit (the fruit will not ripen if it is harvested raw) so that the selection of harvest time is one of the determining factors for the quality of dragon fruit. One of the post-harvest handling of dragon fruit is checking the quality of ripeness of dragon fruit products that have been harvested[4]. Inspection and control of dragon fruit maturity is a very important issue in the management of dragon fruit products. The process of determining the level of maturity of dragon fruit is conventional and there are many obstacles too. The determination is usually by paying attention to changes in the color and size of the dragon fruit. This method is considered less effective if the dragon fruit harvest is very large and there is also a lack of manpower during the dragon fruit sorting process. Sometimes a lot of dragon fruit that is not actually harvestable is also sorted, causing a decrease in the quality of the dragon fruit. To overcome this, we use digital image processing techniques and intelligent systems that aim to determine the level of maturity of the dragon fruit automatically so that there is no need to damage the dragon fruit itself.

This research is part of previous research, namely the classification system for the quality sorting of white dragon fruit (*Hylocereus undatus*) and red dragon fruit (*Hylocereus costaricensis*). Previously, researchers only focused on the classification of dragon fruit quality based on size, so in this study the focus was on the classification of dragon fruit maturity levels which aimed to improve the quality of dragon fruit. There are four classes as system output namely green dragon fruit, red-green dragon fruit, red-yellow dragon fruit and red dragon fruit

## II. RELATED WORKS (OPTIONAL) – 11 PT

Several other studies that serve as references for dragon fruit maturity, namely the design of a dragon fruit ripeness quality sorting tool using the segmentation method based on the HSV color space[4]. Next, identify the hylocereus image using the Discrate Cosine Transform method. In this study, the stages of color conversion from the RGB color space to the grayscale color space were carried out, then the next stage was edge detection. The system is capable of determining the ripeness of dragon fruit with 80% accuracy [5]. The implementation of the backpropagation technique for determining the ripeness of red dragon fruit uses three ripeness levels of dragon fruit: fresh, ripe, and overripe. The achieved system accuracy is 96.67% [6]. The K-Nearest Neighbor method is also used to determine the ripeness of red dragon fruit. The input parameters used are a combination of RGB color characteristics and statistical characteristics (mean), so the accuracy of the system is 93.3%[7]. Another intelligent system used for classification is Naïve Bayes where the method uses a simple probability calculation but the method is quite reliable, especially on numerical data[8]. The Naive Bayes classifier has also been used for classification in data mining [9], identifying microscopic images of acute lymphoblastic leukemia (ALL) with 80% accuracy [8] and classified fungal species according to first-order statistical characteristics with the highest accuracy rate of 98.75%[10]. Based on the results of the above explanation, Naïve Bayes can be used in classifying with high system accuracy.

This study is part of a previous study, a classification system for quality grading of white dragon fruit (Hylocereus undatus) using backpropagation with a system accuracy of 86.67% [11]. We then compared several classification methods such as K-Nearest Neighbor (KNN) and Naive Bayes methods to classify the quality of red dragon fruit (Hylocereus costaricensis) and obtained a system accuracy of 85.33% (KNN). and 86.67% (Naive Bayes) [12]. The focus of this study is the color classification of red dragon fruit (Hylocereus costaricensis) ripeness using the Naïve Bayes algorithm.

## III. METHODS

In the study, several stages were performed to determine the ripeness of red dragon fruit, consisting of image data acquisition of red dragon fruit, digital image processing techniques, and a classification process, as shown in Figure 1.

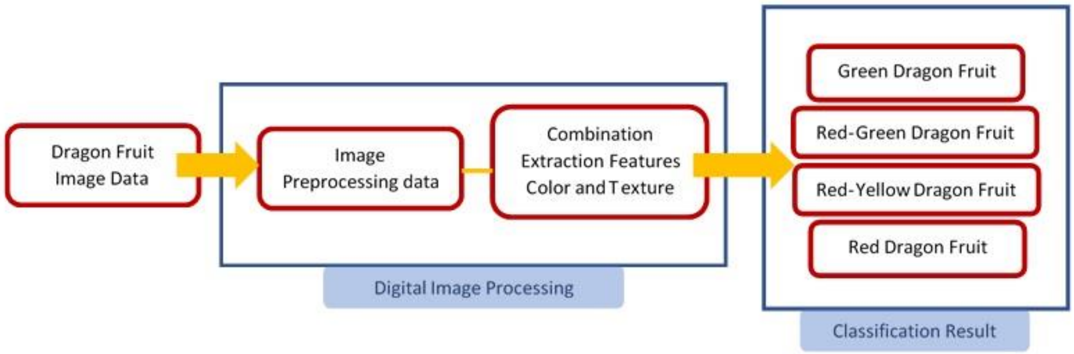


**Fig. 1** The Stages of Research

## *A. Dragon Fruit Image Data*

Srono sub-district is also one of the sub-districts in Banyuwangi Regency which has a fairly large amount of dragon fruit production, which is 2235.5 tons in 2021, especially in Sumbersari village of 657.5 tons in 2020[13]. The process of taking red dragon fruit image data begins with placing the red dragon fruit in a mini studio box. Then the researcher added 3 lights to the mini studio box. It aims to increase the intensity of light so that the resulting image is free from shadows. The next step is to take an image from a distance of 20 cm using a smartphone camera with 13 MP camera resolution, this step is shown in Figure 2.

**Fig. 2** The Process of Retrieving Red Dragon Fruit Image Data

## *B. Digital Image Processing*

This stage consists of two processes, namely preprocessing and feature extraction. The preprocessing process is by cropping the image from the size of 4128 x 3096 pixels to 2301 x 2301 pixels. It aims to reduce the size of the image so as to speed up the computational process. The color variations of the red dragon fruit samples used were green dragon fruit, green-red dragon fruit, yellow-red dragon fruit and red dragon fruit then are shown in Figure 3.

(a) (b) (c) (d)

**Fig. 3** (a) green dragon fruit, (b) green-red dragon fruit, (c) yellow-red dragon fruit and (d) red dragon fruit

### *C. Combination Extraction Features Color and Texture*

Parameters that distinguish the maturity level of dragon fruit based on the color of the dragon fruit, so to take its unique characteristics we use color and texture feature extraction. Find the characteristic values for the red, green, and blue components of each dragon fruit image to get the color characteristic. The RGB value can be used as a comparison parameter because the average RGB value of each object can be different from each other as shown in Figure 4. The RGB color model is expressed in the form of an RGB color index with the following formula :

$$Index\ Red\ Color = \frac{R}{(R+G+B)} \quad (1)$$

$$Index\ Green\ Color = \frac{G}{(R+G+B)} \quad (2)$$

$$Index\ Blue\ Color = \frac{B}{(R+G+B)} \quad (3)$$

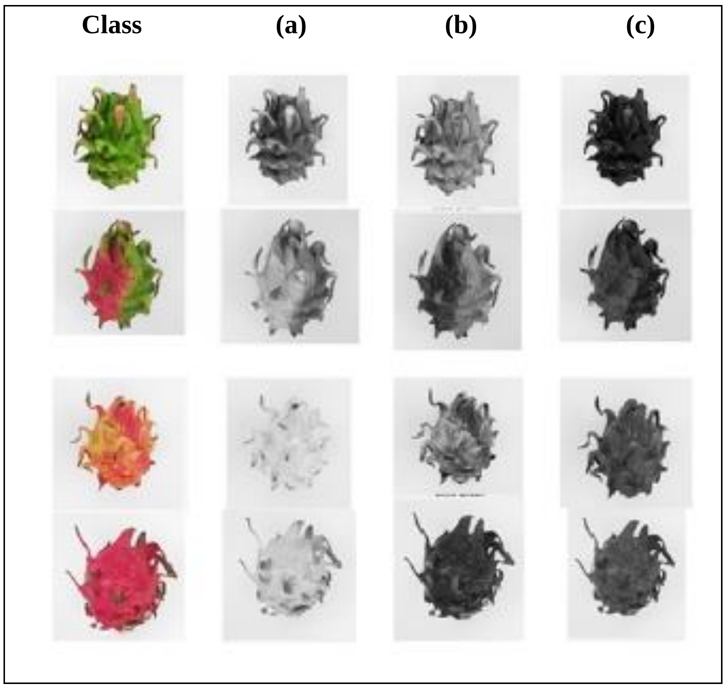


**Fig. 4** Image results for splitting components (a) red, (b) green and (c) blue

Figure 4 shows that the image of the green component has a significant difference for all classes. Then the image of the green component is used as input in finding the texture value in each class. The texture feature extraction used is the feature of the Gray Level Co-Occurrence Matrix (GLCM). GLCM has a square matrix NxN values where N represents the number of gray levels in the image.

The element p(i,j,d,θ) of the GLCM image represents the relative frequency, where I with distance d and direction from position (x,y). The distance (d) used is usually 1 pixel, and the angular direction used is usually 0°, 45°, 90° and 135° [14].The first step to calculate the GLCM feature is to convert the RGB image into a scaled image (greyscale image) then determine the pixel composition and do the number of pixel pairs then create a co-occurance matrix with the transpose matrix. The next step is to normalize the symmetric matrix by calculating the probability of each matrix element so that the GLCM matrix is obtained [15]. The mathematical formulation of texture features is described by the following equations.

$$ASM = \sum_{i=1}^{L}\sum_{j=1}^{L}(GLCM(i,j))^2 \tag{4}$$

$$Contrast = \sum_{i}^{L}\sum_{j}^{L}|i-j|^2 GLCM(i,j) \tag{5}$$

$$IDM = \sum_{i}^{L}\sum_{j}^{L}\frac{(GLCM(i,j))^2}{1+(i-j)^2} \tag{6}$$

$$entropy = -\sum_{i=1}^{L}\sum_{j=1}^{L}\big(GLCM(i,j)\big)\log\big(GLCM(i,j)\big) \tag{7}$$

$$correlation = \sum_{i}^{L}\sum_{j}^{L}\frac{(i-\mu i')(j-\mu j')(GLCM(i,j))}{\sigma i \sigma j} \tag{8}$$

*D. Combination Extraction Features Color and Texture*

Naive Bayes classifiers are classifications based on the conditional probabilities of random variables (class variables) by providing known observations about the values of another set of random variables (characteristic variables) and are common in probability and statistics used for purposes[10]. The advantage of using the Naive Bayes method is that it is relatively easy to implement, efficient to train, can work with binary or polynomial data, and is assumed to be independent, so the method can be implemented on a wide range of datasets is [16]. The formula for the Naïve Bayes theorem is as follows :

$$P(Y|X) = \frac{P(Y)\prod_{i=1}^{q}P(X_i|Y)}{P(X)} \tag{9}$$

$P(Y|X)$ : probability of data with vector X in class Y

$P(Y)$ : initial probability class Y (prior probability)

$\prod_{i=1}^{q}P(X_i|Y)$ : class Y independent probability of all features in vector X

$P(X)$ : probability of X

## IV. RESULTS AND DISCUSSIONS

In this research, the very big influence is the digital image processing process, especially in the extraction of features on the dragon fruit image. Based on the results of Figure 4, the color feature used in this research is the green color feature. This is because the green color feature can represent the colors of the four classes. Then the green color feature is also used as input for the texture feature extraction process using GLCM. Some of the features used are ASM, Contrast, IDM, Entropy and Correlation. The five features are taken with 4 angles, namely 0°, 45°, 90° and 135°. The results of GLCM trait parameters among dragon fruit classes at 0° angle are shown in Table 1. This table shows that only contrast and entropy values can be used for the classification process. This is because the ASM, IDM, and correlation feature values cannot be used because some classes have values that are close to (similar) to other classes on these features. For example, at an angle of 0°, the ASM feature in the Green class has a value of 0.435745, while the Green_Red class has a value of 0.467155 and the Yellow_Red class has a value of 0.457575. The similarity of these values will have an impact on errors in the classification process.

Table 1. The results of the GLCM feature in each Class at an angle of 0°

| Feature | Color class of dragon fruit maturity level | | | |
|---|---|---|---|---|
| | ***Green*** | ***Green-Red*** | ***Yellow-Red*** | ***Red*** |
| ASM 0° | 0,435745 | 0,467155 | 0,457575 | 0,612382 |
| Contrast 0° | 18,75906 | 16,07771 | 12,00043 | 11,84269 |
| IDM 0° | 0,758708 | 0,787144 | 0,81297 | 0,855201 |
| Entropy 0° | 3,121656 | 2,915454 | 2,920368 | 2,05192 |
| Correlation 0° | 0,000571 | 0,000695 | 0,001007 | 0,002156 |

The total parameters used are 9 parameters. These parameters are used as input for the Naive Bayes method. Naive Bayes is a machine learning technique that uses probability calculations and simple statistics. The selection of Naïve Bayes as a method for classifying is because each attribute used for classification will contribute to decision making with the weights of each attribute. The weight of each attribute has an equal importance in the classification process. The number of training data used are 225 images, consisting of 53 green data, 46 green-red data, 63 yellow-red data and 63 red data. While the testing data used were 95 images, consisting of 22 green data, 19 green-red data, 27 yellow-red data and 27 red data. In the training process, the method of calculating cross validation is used. Cross validation is a statistical method that can be used to evaluate the performance of a model or algorithm where the system randomly divides data into two data, namely training and testing data. The percentage used is 70% the amount of data as training data and 30% as testing data. To determine the level of system accuracy is calculated using the calculation of the confusion matrix. The system testing is display as shown in Figure 5.

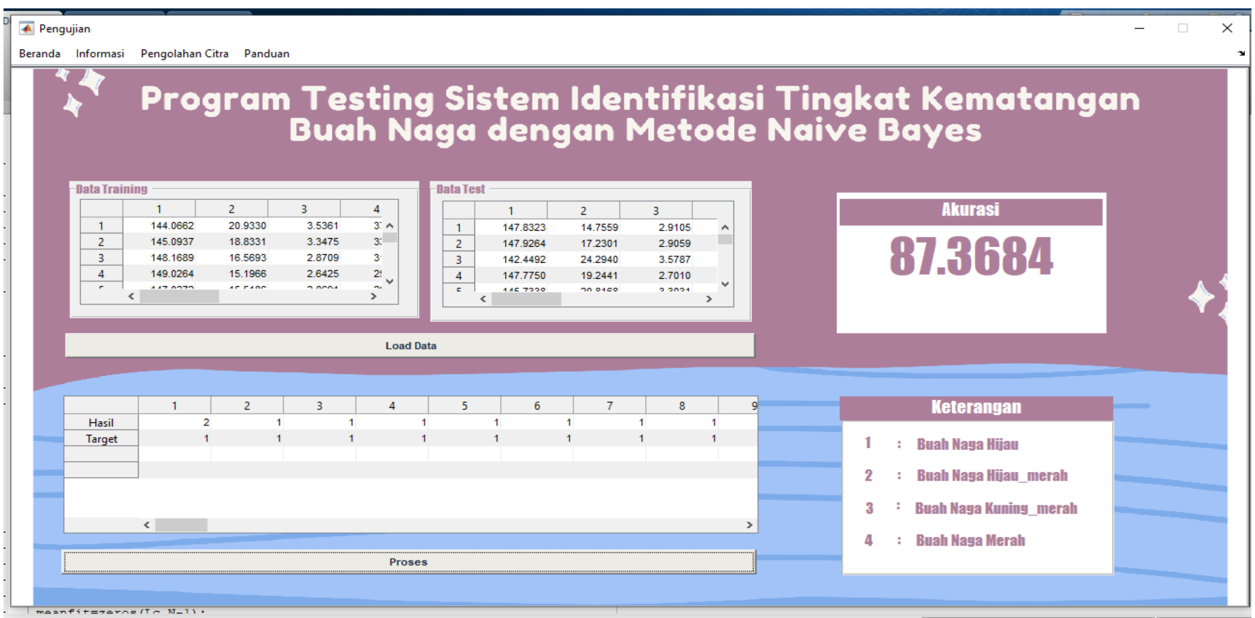


Fig. 5 Display application results of system testing

Figure 5 shows that we have 70% of the total training data and 30% of the test data, the Naive Bayes method performs the classification process, and the results are displayed in the Excel viewer. Each class is given a code, namely code 1 for green, code 2 for red and green, code 3 for yellow red and code 4 for red. Based on this figure, the Naive Bayes method can classify the ripeness of dragon fruit with 87.3684% or 87.37% accuracy.

## V. CONCLUSIONS AND RECOMMENDATIONS

Based on the results of this study, it can be concluded that image processing techniques, especially the feature selection process, greatly affect the level of accuracy of the Naïve Bayes system in classifying images of dragon fruit maturity levels into 4 classes, namely green class, red-green class, yellow-red class and red class. The accuracy of the system determined using test data according to the Naive Bayes method is 87.37%. The Naive Bayes classification accuracy values in this study did not reach 100%. This is because the feature extraction result values for a particular class have similar values, which complicates the classification process. To improve the results, another classification method is needed as a comparison to obtain an effective and accurate method for determining the ripeness of red dragon fruit (Hylocereus costaricensis).

## VI. ACKNOWLEDGMENTS (OPTIONAL)

We would like to thank the Center for Research and Community Service (P3M) at Politeknik Negeri Jember for supporting and funding this study.

## VII. REFERENCES

[1] F. Maulida, W. D. Widodo, dan K. Suketi, “Kriteria Kematangan Pascapanen Buah Naga Merah (Hylocereus polyrhizus (Weber) Britton & Rose) dengan Umur Panen yang Berbeda.,” IPB University, 2018.

[2] D. P. K. Banyuwangi, “Data Tahunan Bidang Perkebunan dan Hortikultura,” *Dinas Pertanian Kabupaten Banyuwangi*, 2018. https://dinaspertanian.banyuwangikab.go.id (diakses Agu 15, 2022).

[3] D. Fitrada, “APLIKASI IMAGE PROCESSING UNTUK MENENTUKAN TINGKAT MUTU BUAH NAGA (Hylocereus undatus) SECARA NON-DESTRUCTIVE,” INSTITUT PERTANIAN BOGOR, 2010.

[4] Sustiono dan W. S. Pambudi, “Rancang Bangun Alat Pemilihan Kualitas Kematangan Buah Naga Menggunakan TeknikImage Processing dengan Metode Image Segmentation HSV,” *J. Sains dan Inform.*, vol. 1, no. 2, hal. 28–37, 2015.

[5] R. Wulanningrum dan N. V. W. Lelitavistara, “Discrete Cosine Transform Untuk Identifikasi Citra Hylocereus Costaricensis,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 6, no. 2, hal. 353, 2015, doi: 10.24176/simet.v6i2.472.

[6] R. D. M. Bimantaka, “IDENTIFIKASI KEMATANGAN BUAH NAGA MERAH MENGGUNAKAN METODE BACKPROPAGATION BERBASIS ANDROID,” Universitas Teknologi Yogyakarta, Yogyakarta, 2017.

[7] F. Nisa’, “Sistem Identifikasi Citra Hylocereus costaricensis dengan Menggunakan Metode K-Nearest Neighbor dan Mean Statistik,” Universitas Nusantara PGRI Kediri, Kediri, 2018.

[8] S. Shidada *et al.*, “Identifikasi Acute Lymphoblastic Leukemia pada Citra Mikroskopis Menggunakan Algoritma Naïve Bayes Evaluas i,” vol. 14, no. 1, hal. 78–83, 2021.

[9] D. Ariadi dan K. Fithriasari, “Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer,” *J. SAINS DAN SENI ITS Vol. 4, No.2*, vol. 4, no. 2, hal. 248–253, 2015.

[10] A. A. Mahran, R. K. Hapsari, dan H. Nugroho, “PENERAPAN NAIVE BAYES GAUSSIAN PADA KLASIFIKASI JENIS JAMUR BERDASARKAN CIRI STATISTIK ORDE PERTAMA,” *J. Ilm. NERO*, vol. 5, no. 2, hal. 91–99, 2020.

[11] Z. E. Fitri, A. Baskara, M. Silvia, A. Madjid, dan A. M. N. Imron, “Application of Backpropagation Method for Quality Sorting Classification System on White Dragon Fruit (Hylocereus undatus),” in *IOP Conference Series: Earth and Environmental Science*, Apr 2021, vol. 672, no. 1, doi: 10.1088/1755-1315/672/1/011001.

[12] Z. E. Fitri, A. Baskara, A. Madjid, dan A. M. N. Imron, “Comparison of Classification for Grading Red Dragon Fruit (*Hylocereus Costaricensis*),” *J. Nas. Tek. Elektro*, vol. 11, no. 1, hal. 43–49, 2022, doi: 10.25077/jnte.v11n1.899.2022.

[13] B. K. Banyuwangi, *Kecamatan Srono Dalam Angka 2021*. Banyuwangi: CV. Anugerah Setia Abadi, 2022.

[14] R. Enggar Pawening, W. Ja, dan F. Shudiq, “Klasifikasi Kualitas Jeruk Lokal Berdasarkan Tekstur dan Bentuk Menggunakan Metode k-Nearest Neighbor (k-NN),” *J. Ilmu Komput. dan Desain Komun. Vis.*, vol. 1, no. 1, hal. 10–17, 2020.

[15] Z. E. Fitri, U. Nuhanatika, A. Madjid, dan A. M. N. Imron, “Penentuan Tingkat Kematangan Cabe Rawit (Capsicum frutescens L.) Berdasarkan Gray Level Co-Occurrence Matrix,” *J. Teknol. Inf. dan Terap.*, vol. 7, no.

1, hal. 1–5, 2020, doi: 10.25047/jtit.v7i1.121.
[16] A. Ghofur, "Implementasi Metode Klasifikasi Naive Bayes Untuk Memprediksi Kualitas Cabai," *J. Ilm. Inform.*, vol. 1, no. 1, hal. 32–38, 2016, doi: 10.35316/jimi.v1i1.441.