Determination Characteristic and Classification the Types of Orange Using UV-Vis Spectrophotometer by K-Nearest Neighbor Algorithm

Abel Harditio Pratama, Anak Agung Ngurah Gunawan*, Hery Suyanto

Department of Physics, University of Udayana at Bali, Denpasar 80119, Indonesia

Corresponding Author Email: a.a.n.gunawan.unud@gmail.com

ABSTRACT

It has been researched at Laboratorioum Terpadu Faculty of Science and Mathematics. The purpose of this study is determining to distinguish in characteristic between sweet orange and sour orange, it makes sure wavelength and absorbance using UV-Vis Spectrophotometer. Furthermore, it is also determining classification between sweet orange and sour orange by predicting between actual data and predictive data. The method used is K-Nearest Neighbor algorithm using MATLAB 2015. Sweet orange and sour orange are researched for their characteristic to be analyzed whether they have significant differences in wavelength and absorbance. The KNN algorithm is used to see its ability in classification to predict sweet orange and sour orange data from actual data. Through this study, it is found that resulting characteristic of sweet oranges and acid oranges has a different, it can be seen based on forms wavelength and absorbance. Sweet oranges and acid oranges have they own characteristic. On the classification, prediction data results showing that 23 correct data from 40 actual data obtained a percentage of 67.5 % by using K-Nearest Neighbor. The finding of this research may serve as reference. The finding of this study may serve as reference for other researchers to be developed further.

1. INTRODUCTION

Basically, research is process of inquiry or search for something (facts and principles) that is done systematically, carefully, and critically. From this understanding, it can be concluded that research is a method of finding truth, so research is a method of thinking critically.

Agriculture sector is the mainstay sector in national economic development. Its assignment included contributing to providing foreign exchange through exports, provision of food and industrial raw materials, and providing employment.

Oranges (*Citrus sp*) is the of most popular world fruits. It also contains important nutritional elements for health. Citrus is good source of vitamin C (ascorbic acid), phenolic compound, flavonoid, folic acid, potassium, pectin and antioxidant properties [1].

Originally invented for estimating vitamin content in US military rations, spectrophotometer later became one of the most widely used measuring instruments of all time in various fields of experimental science. As Nobel laureate chemist Bruce Merrifield vowed for Spectrophotometer of being "the most important instrument ever developed toward the advancement of bioscience". The instrument is commonly used in chemical education, biochemistry, chemical physics and material science [2, 3].

For two decades, researchers and practitioners have been using UV/Vis spectrophotometers to estimate from absorbances at several wavelength. The accuracy and robustness require a local calibration. Taking into account the local specifications of the water matrices, samples are collected, measured with the special device and concentration are measured with laboratory analysis [4, 5].

K-Nearest Neighbor has been one of the most popular classification algorithms. The classical KNN algorithm is based on calculating the distance between the test data instances to be classified and all of the instances in the training data set and finding the closest K number of training instances. After detecting the K number of closest training instances, the KNN algorithm applies majority voting which is the process of detecting the data class with the maximum number of instances among the K selected instances [6].

Since the classical KNN algorithm is completely based on individual instance proximities, it heavily suffers from high computation cost. In addition, since the algorithm decision-making strategy is relying on the individual instance proximities rather than stronger class representations, the algorithm's classification accuracy is also not adequate for modern data big analysis that requires rapid and accurate classification results [6].

In this case, it will be researched which some oranges of lumajang are being used as samples. It determines to distinguish between sweet oranges and acid oranges, both characteristics and classification. For determination characteristics is used UV-Vis spectrophotometer, which it results wavelength and abosrbansi. Meanwhile determination classification is used K-Nearest Neighbor methods by Matlab 2015 software. Furthermore, the conclusion is determination characteristic and classification a few of oranges of kintamani using UV-Vis spectrophotometer by K-Nearest Neighbor.

2. THEORETICAL

2.1 Orange

Orange (*Citrus sp*) is a plant from Asian. China is believed to be the place where oranges first time grew in both the tropics and sub-tropics. Once upon a time it was around one hundred years ago, oranges have been grown in Indonesia. Indonesia is tropics where it is variety of oranges may be found a whole Indonesian island. Even there are several oranges have been favoring in both region and national [7].

Orange is one of the horticultural commodities that serves as nutrition, a source of income, and a source of foreign exchange. The contribution of citrus agro-industry in increasing income will give the development of the oranges [7].

2.2 UV-Vis spectrophotometer

Spectrophotometer is a device that characterizes chemical substances in terms of their capability of absorbing different parts of electromagnetic spectrum. There are different ranges of wavelengths those are covered by different types of spectrophotometers, e.g. IR spectrophotometer, visible light spectrophotometer, UV-Vis spectrophotometer etc., [8-9]. Since the spectrophotometric method uses the phenomenon of absorbance of EM spectrum, this is also called Absorption Spectrophotometry. In academia, this is commonly taught topic in introductory undergraduate chemistry, pharmacy, material catherization, and electronic materials courses [10, 11].

Figure 1. UV-Vis spectrophotometer

Spectrophotometry is the study of the use of spectrophotometers. Spectrophotometer is a device consisting of a spectro and a photometer. Spectrophotometer produces a ray from spectrum by a certain wavelength.

If radiation is passed through a colored solution, then the radiation by certain wavelength will be absorbed selectively and other radiation will be transmitted. Absorbance is the ratio of the intensity of the absorbed light with the intensity of the light is coming.

2.3 UV-Vis spectrophotometer instrumentation

Instrumentation for UV-Vis spectrophotometer consists of five components.
a. Light source
The light source for the spectrum is the argon lamp on UV-Vacuum, deuterium lamp or hydrogen lamp on Ultraviolet Spectrophotometry, xenon lamp and wolfram lamp on UV-Vis Spectrophotometer.
b. Sample place
Sample place usually can be used for put down the cuvette. All liquid objects will be filled in cuvette and then the cuvette is put down into sample place of UV-Vis Spectrophotometer.
c. Monochromator
Monochromator is a device used to produce a beam of radiation with one wavelength. Monochromator usually consist of slit, lens, mirror, and prism.
d. Detector
There are two types of detector, photon detector and heat detector.
e. Recorder
Electrical signal which received by detector, it will be recorded as spectrum

2.4 K-nearest neighbor algorithm

In 1968, Cover and Hart proposed an algorithm the K-Nearest Neighbor, which was finalized after some time. K-Nearest Neighbor can be calculated by calculating Euclidean distance, although other measures are also available but through Euclidean distance we have splendid intermingle of ease, efficiency and productivity [12].

Algorithm the K-Nearest Neighbor (KNN) is a method to classify about object based on training data which its nearest distance by the object. KNN included supervised learning algorithm which its result of the new query instance is classified based on the majority from the KNN category. The class that appears the most it will be the classification class. The purpose of this algorithm is to classify the new objects based on attributes and training data [13, 14].
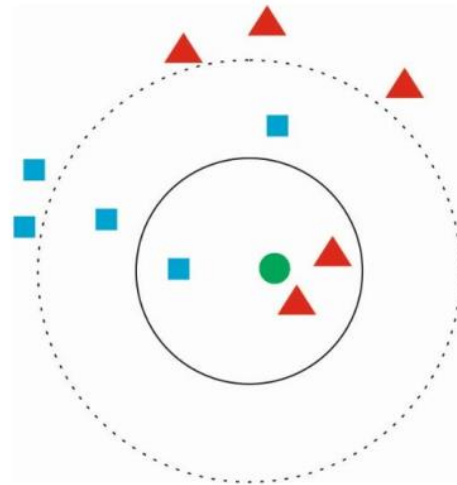

Figure 2. K-Nearest neighbor illustration

The figure above contains two class classifications illustrated as the "blue square" class dan the "red triangle" class. And then put in a value test as "green circle". The KNN method shows the result that "green circle" is more towards "red triangle". At a certain range obtained by the nearest neighbor that is "red triangle" of two pieces and "blue square" there is only one piece. The nearest neighbour will be calculated based on the Euclidean distance in the following equation.

$$d = \sqrt{\sum_{i=1}^{n}(x_{training}^1 - x_{testing})^2} \qquad (1)$$

There is an example of applying the KNN algorithm. Two variables are acid resistance and strength, it finds out classification for determination between good quality or bad quality on tissue. The following are four training data.

Table 1. Training data

X1 = Acid Resistance	X2 = Strength	Classification
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good

As a case, for example a factory has produced a new system that has passed laboratory testing, which each of the sample is 3 as acid resistance and 7 as strength. To predict classification the new system and then it will be calculated using K-Nearest Neighbour algorithm.

As for the steps to count the KNN as follows.

(1) Determine the parameter k, for example k=3.

(2) Calculate distance between training data and testing data. Training data will be calculated proximity has coordinates (3 and 7).

Table 2. Distance calculation

X1=Acid Resistance	X2=Strength	The Range Quadrate with the New Data
7	7	$(7-3)^2+7-7^2=15$
7	4	$(7-3)^2+4-7^2=25$
3	4	$(3-3)^2+4-7^2=9$
1	4	$(1-3)^2+4-7^2=13$

(3) Sort distance and determine K-Nearest Neighbour.

Table 3. Determination K-Nearest neighbour

X1=Acid Resistance	X2= Strength	New Data	Minimum	Including k=3
7	7	15	3	Yes
7	4	25	4	No
3	4	9	1	Yes
1	4	13	2	Yes

(4) Collecting category Y from the nearest neighbour row. On the second row does not including to category Y due to the data exceed k-nearest = 3.

Table 4. Determination category Y nearest neighbor

X1 = Acid Resistance	X2 = Strength	New Data	Minimum	Including k = 3	KNN
7	7	15	3	Yes	Bad
7	4	25	4	No	-
3	4	9	1	Yes	Good
1	4	13	2	Yes	Good

Based on table above obtained two good quality on tissue and a bad quality on tissue. The nearest neighbour obtained is more dominate the good quality. The conclusion shows that tissue paper that just passed laboratory test with 3 and 7 as coordinate point is including good categories.

2.5 MATLAB

MATLAB (Matrix Laboratory) is a program for analysis and numerical computation, it is an advanced mathematical programming language that is designed with the rationale and uses the properties and forms of the matrix [15].

MATLAB has developed into a programming environment that contains built-in functions to perform processing task, linier algebra, and etc. MATLAB also contains toolbox as additional function specially for applications [16].
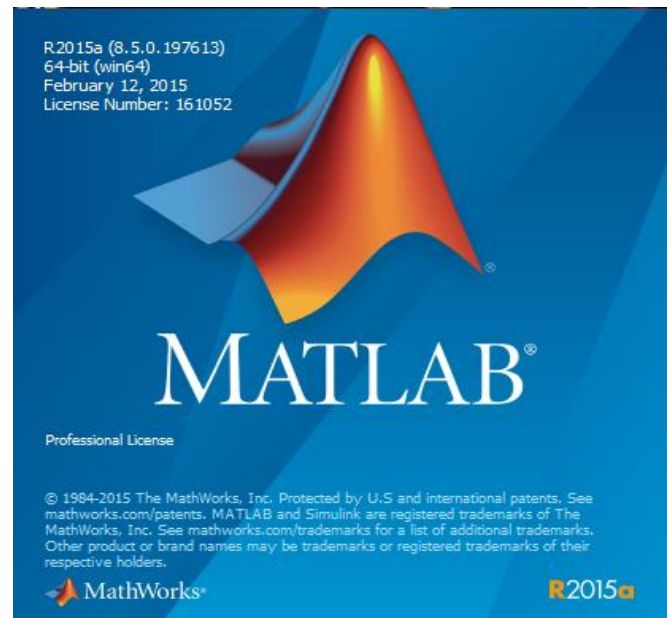

Figure 3. MATLAB software

3. METHODS

3.1 Location and time

The research will be held At Laboratorioum Penelitian Terpadu Fakultas Matematika dan Ilmu Pengetahuan Alam. It will be going for two months start from August until September 2019.

3.2 Data collection

The data used in this research is obtained through the result of wavelength and absorbance data. They are obtained through UV-Vis Spectrophotometer based on sweet oranges and acid oranges which they are ten pieces each. It is determination the difference characteristics and classification in sweet oranges and acid oranges especially lumajang oranges.

3.3 Data processing

In data processing for classification, the process of grouping data to determine the variable to be used, performing data representation into numerical form and doing data sharing into training data and testing data
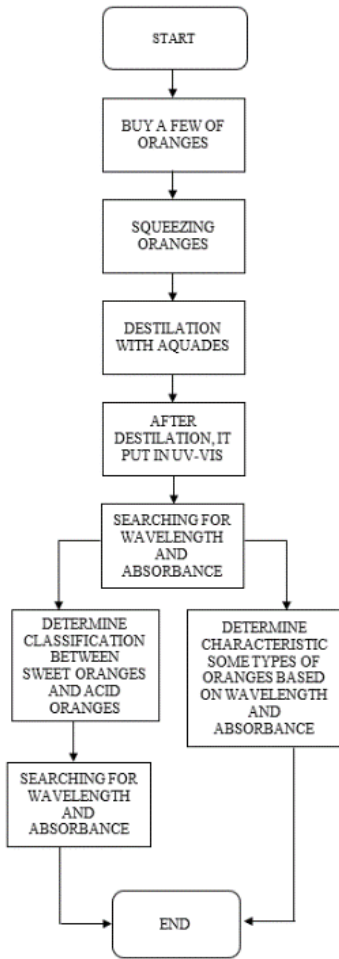
3.4 Research flowchart

Figure 4. Research flowchart

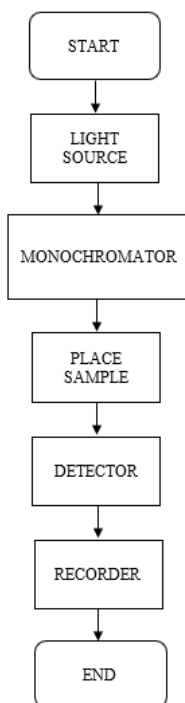## 3.5 UV-Vis spectrophotometer flowchart

Figure 5. UV-Vis spectrophotometer instrumentation flowchart

3.6 Characteristic and classification

In this research it will determine to distinguish of characteristic between sweet orange and acid orange based on wavelength and absorbance. Furthermore for classification, it predicts two variables are actual data and prediction data based on sweet orange and acid orange. These oranges were bought at Traditional Market in Kintamani, Bali. The type of oranges used is Lumajang orange.

3.7 Program design

Some researchers have been doing to classify of data. And then, in this research will be used K-Nearest Neighbour method. Graphical user interface design can be seen on Figure 2.
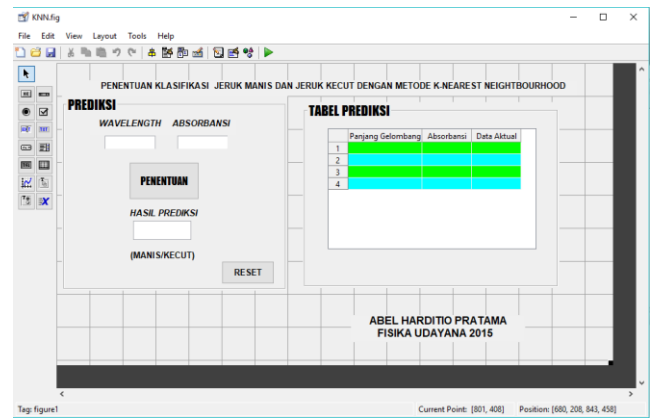

Figure 6. Program design

4. RESULT AND DISCUSSION

4.1 Determination wavelength and absorbance

The result of determination wavelength and absorbance can be seen on Figure 6.

On Figure 7 it appears that absorbance increases with increasing wavelength and reaching absorbance value to wavelength at point 2 is 266.00 nm.

While on the Figure 8 it appears that absorbance does not increase with increasing wavelength and reaching absorbance value to wavelength at point 2 is 272.00 nm.
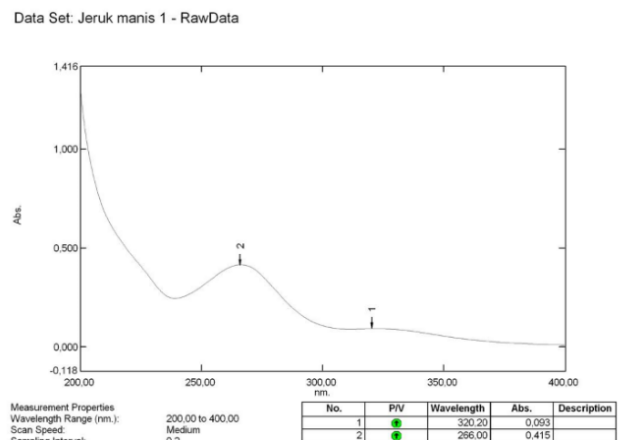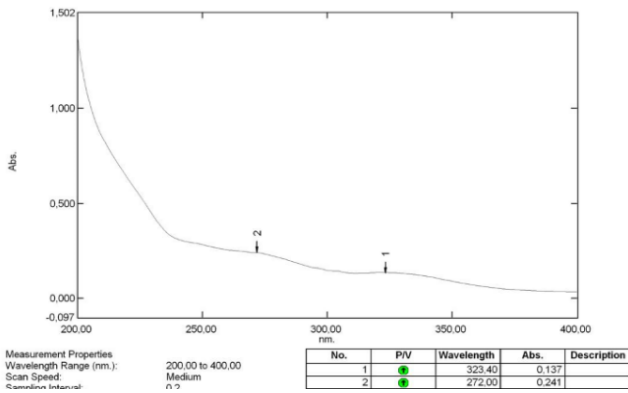

Figure 7. Wavelength and absorbance charts in sweet oranges 1

Data Set: Jeruk Kecut 1 - RawData

Measurement Properties
Wavelength Range (nm.): 200,00 to 400,00
Scan Speed: Medium
Sampling Interval: 0,2

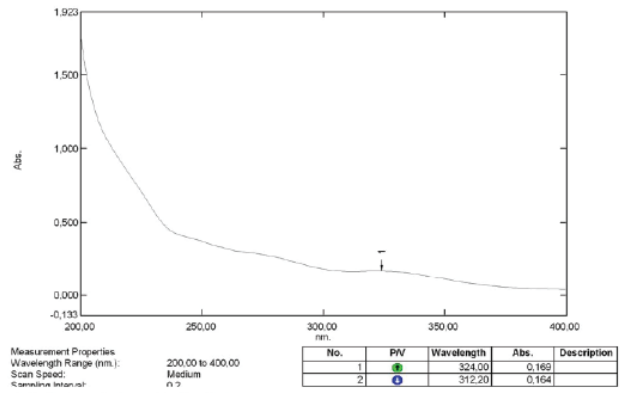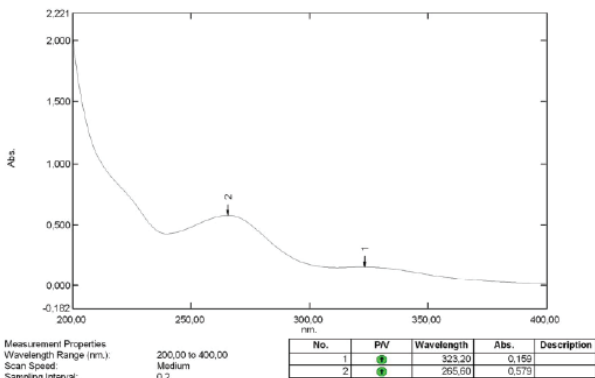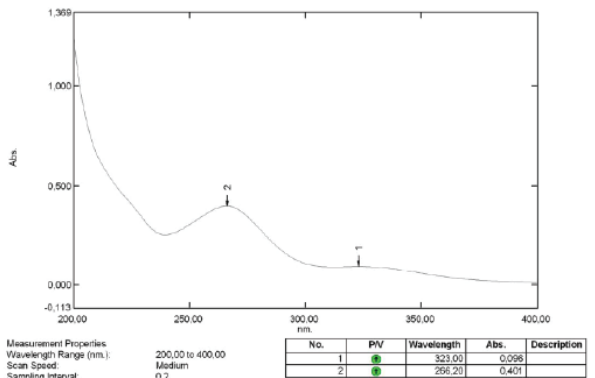| No.	P/V	Wavelength	Abs.	Description
1		323,40	0,137	
2		272,00	0,241	

Figure 8. Wavelength and absorbance charts in sweet oranges 2

4.2 Analysis of differences in characteristics of oranges

Data Set: Jeruk manis 2 - RawData

Measurement Properties
Wavelength Range (nm.): 200,00 to 400,00
Scan Speed: Medium
Sampling Interval: 0,2

No.	P/V	Wavelength	Abs.	Description
1		323,20	0,159	
2		265,60	0,579	

Data Set: Jeruk manis 3 - RawData

Measurement Properties
Wavelength Range (nm.): 200,00 to 400,00
Scan Speed: Medium
Sampling Interval: 0,2

No.	P/V	Wavelength	Abs.	Description
1		323,00	0,096	
2		266,20	0,401	

Data Set: Jeruk manis 4 - RawData

Measurement Properties
Wavelength Range (nm.): 200,00 to 400,00
Scan Speed: Medium
Sampling Interval: 0,2

No.	P/V	Wavelength	Abs.	Description
1		323,40	0,070	
2		265,00	0,301	

Figure 9. Characteristic of sweet oranges

Data Set: Jeruk Kecut 2 - RawData

Measurement Properties
Wavelength Range (nm.): 200,00 to 400,00
Scan Speed: Medium
Sampling Interval: 0,2

No.	P/V	Wavelength	Abs.	Description
1		324,00	0,169	
2		312,20	0,164	

Data Set: Jeruk Kecut 3 - RawData

Measurement Properties
Wavelength Range (nm.): 200,00 to 400,00
Scan Speed: Medium
Sampling Interval: 0,2

No.	P/V	Wavelength	Abs.	Description
1		323,80	0,106	
2		270,80	0,206	

Data Set: Jeruk Kecut 4 - RawData

Measurement Properties
Wavelength Range (nm.): 200,00 to 400,00
Scan Speed: Medium
Sampling Interval: 0,2

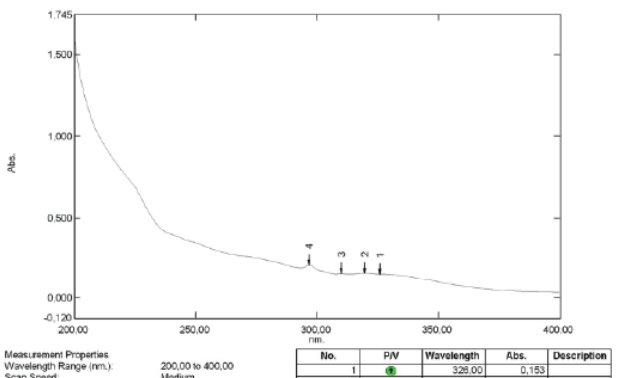| No.	P/V	Wavelength	Abs.	Description
1		326,00	0,153	
2		319,80	0,159	

Figure 10. Characteristic of acid oranges

In section 5.1 it has been explained that there are different forms wavelength and absorbance between sweet orange and acid orange. Each orange has been proven to have a significant difference which sweet orang has hill waves while acid orange has not, and it tends to be more horizontal. As evidence there are some samples between sweet oranges and acid oranges that have different form wavelength and absorbance can be seen on the Figures above.

On the Figure 8, there are three different forms wavelength and absorbance. Which at point 2 it can be seen that they have hill waves, it is starting from *jeruk manis 3, jeruk manis 4* and *jeruk manis 5*. Each of samples has point 2 is different with wavelength value and absorbance value.

On the Figure 9, there are three different forms wavelength and absorbance. There are no hill waves, it is starting from *jeruk kecut 3, jeruk kecut 4* and *jeruk kecut 5*. Each of samples has point 2 is different with wavelength value and absorbance value.

4.3 Classification sweet oranges and acid oranges

On the classification, it will determine to predict between sweet oranges and acid oranges using K-Nearest Neighbour algorithm method based on wavelength value and absorbance value. Software used for predicting taste of oranges is MATLAB 2015, and the Graphical User Interface has been designed the following in section 3.7.

Table 5. Classification of sweet oranges

Wavelength	Absorbance	Actual Data	Prediction Data	Information
320.2	0.093	1	1	True
266	0.415	1	1	True
323.2	0.159	1	0	False
265.6	0.579	1	1	True
323	0.096	1	0	False
266.2	0.401	1	1	True
323.4	0.07	1	0	False
265	0.301	1	1	True
324.2	0.069	1	0	False
265.6	0.315	1	1	True
265.6	0.0057	1	0	False
398.8	0.01	1	0	False
324	0.128	1	0	False
266.4	0.459	1	1	True
323.2	0.129	1	0	False
266	0.461	1	1	True
324	0.136	1	0	False
264.2	0.522	1	1	True
323.4	0.107	1	0	False
265	0.427	1	1	True

Table 6. Classification of acid oranges

Wavelength	Absorbance	Actual Data	Prediction Data	Information
323.4	0.0137	0	0	True
272	0.241	0	1	False
324	0.169	0	0	True
312.2	0.164	0	0	True
323.8	0.106	0	0	True
270.8	0.206	0	0	True
326	0.153	0	0	True
319.8	0.159	0	0	True
322.8	0,142	0	0	True
310.2	0.138	0	0	True
327.2	0.131	0	0	True
323.6	0.131	0	0	True
323.6	0.182	0	0	True
309.8	0.174	0	0	True
318.2	0.182	0	1	False
262.6	0.322	0	0	True
323.8	0.318	0	0	True
269.6	0.237	0	1	False
323.2	0.163	0	0	True
305.2	0.17	0	0	True

The Table 6 shows the prediction results of sweet oranges with the KNN method.

Actual data of this research will be consisted to be two parts are one (1) as sweet taste and zero (0) as acid taste. By using the KNN method, it will results prediction data. However, actual data is one and the prediction data output is one, then the results are accurate and the opposite. And then the actual data is one and the prediction data output is zero, then the results are not accurate. Each sample of sweet oranges and acid oranges will be predicted to determine whether actual data and prediction data will obtain accurate data.

The following tables shows the prediction results of sweet oranges with the KNN method.

Two tables above show the result of predicting by using the KNN method.

There are 40 data on classification of sweet orange and acid orange. Based on the table above, the result of prediction data with the correct amount are 27 data. The result of the percentage of classification by the K-Nearest Neighbor method by 67.5 %.

5. CONCLUSION

The researching has proved that there are different characteristics between sweet oranges and acid oranges based on forms wavelength and absorbance by using UV-Vis

Spectrophotometer. For characteristic of sweet oranges, it has hill waves starting from *jeruk manis 1* untill *jeruk manis 10*. Meanwhile, characteristic of acid oranges has not hill waves starting from *jeruk kecut 1* untill *jeruk kecut 10*. On the classification of sweet oranges and acid oranges, it is 23 correct data from 40 data obtained a percentage of 67.5 %. This study can be conducted further for determining characteristics on other fruits using UV-Vis Spectrophotometer. K-Nearest Neighbor is recommended using for processing data on classification.

ACKNOWLEDGMENT

REFERENCES

[1] Hermin, P.K., Anto, B., Agung, S., Yuriza, E., Annisa, F., Dina, R.P. (2018). The characterization of citrus Sp. from Parang Island Karimun Jawa based on morphological, DNA barcoding and nutritional analysis. International Journal of Genetics and Molecular Biology, 10(3): 26-38. https://doi.org/10.5897/IJGMB2018.0167

[2] Rafi, M., Jannah, R., Heryanto, R., Kautsar, A., Septaningsih, D.A. (2018). UV-Vis spectroscopy and chemometrics as tool for identification and discrimination of four curcuma species. International Food Research Journal, 25(2): 643-648.

[3] Etebu, E., Nwauzoma, A.B. (2014). A review on sweet orange (*Citrus Sinensis* L Obseck): Health, diseases and management. American Journal of Research Communication, 2(2): 33-70.

[4] Lepot, M., Aubin, J.B., Clemens, F.H.L.R., Masics, A. (2017). Outlier detection in UV/Vis spectrophotometric data. Urban Water Journal, 14(9): 908-921. https://doi.org/10.1080/1573062X.2017.1280515

[5] Ashfaque, A., Islam, Md.R., Faria, I.J. (2017). Development and validation of low-cost visible light Spectrophotometer. International Conference on Advances in Electrical Engineering (ICAEE), Dhaka, Bangladesh.

https://doi.org/10.1109/icaee.2017.8255437

[6] Tulgar, T., Haydar, A., Ersan, I. (2018). A distributed K-Nearest neighbour classifier for big data. Balkan Journal of Electrical and Computer Engineering, 6(2): 105-111. https://doi.org/10.17694/bajece.419551

[7] Devy, N.F., Hardiyanto. (2017). Citrus diversity of west sumatera based on morphology, inter-simple sequence repeats and its combined analysis. Biotika, 5(18): 32-39.

[8] Rutuja, S.S., Rajashri, B.P., Pranit P.G. (2015). UV-Visible spectroscopy a review. International Journal of Institutional Pharmacy and Life Sciences, 5(5): 490-505.

[9] Pekamwar, S.S., Kalyankar, T.M., Tembe, B.V., Wadher, S.J. (2015). Validated UV-Visible spectrofotometric method for simulataneous estimation of cefixime and moxifloxacin in pharmaceutical dosage form. Journal of Applied Pharmaceutical Science, 5(1): 037-041. https://doi.org/10.7324/japs.2015.50107

[10] Acharjya, S.K., Mallick, P., Panda, P., Kumar, K.R., Annapurna, M.M. (2011). Spectrophotometric methods for the determination of zolmitripan in bulk and pharmaceutical dosage forms. Journal of Advanced Scientific Research, 2(3): 42-47. https://doi.org/10.4103/0110-5558.72425

[11] Bhawani, S.A., Fong, S.S., Ibrahim, M.N.M. (2015). Spectrophotometric analysis of caffeine. International Journal of Analytical Chemistry. http://dx.doi.org/10.1155/2015/170239

[12] Kataria, A., Singh, M.D. (2013). A review of data classification using K-Nearest neighbor algorithm. International Journal of Emerging Technology and Advanced Engineering, 3(6): 364-369.

[13] Sadegh, B.I., Mohammad, B. (2013). Application of K-Nearest Neightbor (KNN) approach for predicting economic event: Theoretical background. S.B. Imandoust et al. Int. Journal of Engineering Research and Applications, 3(5): 605-610.

[14] Safri, Y.F., Arifudin, R., Muslim, M.A. (2018). K-nearest neighbor and naive Bayes classifier algorithm in determining the classification of healthy card Indonesia giving to the poor. Scientific Journal of Informatics, 5(1): 9-10. https://doi.org/10.15294/sji.v5i1.12057

[15] Purnima, R. (2015). Application of Laplace transforms to solve ODE using MATLAB. Journal of Informatics and Mathematical Science, 7(2): 93-97.

[16] Mahmood, J.R., Selman, N.H. (2016). Four MATLAB-Simulink models of photovoltaic system. International Journal of Energy and Environment, 7(5): 417-426.