

Categorizing Sugarcane Production Based On Factors Affecting Productivity With The K-Nearest Neighbor Algorithm

Anggi Pratiwi¹, A M H Pardede², I Gusti Prahmana³

^{1,2,3} Informasi System, STMIK Kaputama

Jl. Veterans No. 4A-9A, Binjai, North Sumatra, Indonesia

Anggipratiwi2099@gmail.com^{1*}, akimmhp@live.com², igustiprahmana4@gmail.com³

Abstract

Sugarcane (*Saccharum Officinarum*) is an annual plantation crop, which has its own characteristics, because the stem contains sugar. To classify the results of sugarcane production, currently still using the manual method by only looking at the current conditions of sugarcane production. This is less efficient because there is no calculation process in grouping sugarcane. So that mistakes can occur in grouping sugarcane production to get good results or not in the assessment of sugarcane grouping at PTPN II Kwala Madu. For this reason, the author will create an alternative application system that can group sugarcane production with the K-Nearest Neighbor algorithm to find out the best type of sugarcane production based on the factors. The application made by the author uses the PHP programming language and uses the MySQL database as data storage. The system is made as easy as possible to make it easier for users to use and understand later.

Keywords: Clustering, Sugarcane, K-Nearest Neighbor, PHP, Mysql

1. Introduction

Sugarcane (*Saccharum Officinarum* L.) is an annual plantation crop, which has its own characteristics, because the sugar is contained in the stem. Sugarcane belongs to the grass family (graminae) such as rice, glagah, corn, bamboo and others [1]. Sugar cane which in English is called sugar cane has a high sugar content. Therefore, sugar cane is known for its utilization as a staple for sugar production. Sugarcane is a genus of *saccharum* which has the highest sucrose content and the lowest fiber content [2].

Sugar is one of the staple foods consumed by all Indonesian people. The increasing population, people's income and the food and beverage industry, the people's need for sugar is getting higher. Therefore, the plantation industry, including the downstream sugar industry, is the leading sector in increasing employment, economic growth, and encouraging an even distribution of people's income distribution, especially in rural areas. Therefore, to meet the needs of the people, sugar production must be more than consumption.

To classify the results of sugarcane production, currently still using the manual method by only looking at the current conditions of sugarcane production. This is less efficient because there is no calculation process in grouping sugarcane. So that mistakes can occur in grouping sugarcane production to get good results or not in the assessment of sugarcane grouping at PTPN II Kwala Madu.

For this reason, the author will create an alternative application system that can group sugarcane production with the K-Nearest Neighbor algorithm to find out the best type of sugarcane production based on the factors. The application made by the author uses the PHP programming language and uses the MySQL database as data storage. The system is made as easy as possible to make it easier for users to use and understand later.

2. Literature

Many core companies face difficulties in modeling success rates in broiler farming because the large volume of data comes from breeders who partner with companies, making it difficult for companies to determine the success rate of broiler production. Determining the success rate of production is very helpful in selecting breeders who will be guided appropriately, so that it can help make the right decision. To manage the growing amount of data, data mining and the K-Nearest Neighbor (KNN) algorithm are needed in the classification process. This study aims to classify the success rate of chicken production

2.1. K-Nearest Neighbor

K-Nearest Neighbor is a method that uses a supervised algorithm where new testing data is classified based on the majority of classes in KNN. The purpose of this algorithm is to classify new objects based on attributes and training data. The classifier does not use any model to match and is only based on memory. The principle of KNN is to find the k objects from the training data that are closest to the testing data. The KNN algorithm is very simple, works based on the shortest distance from the testing data to the training data to determine the k-nearest neighbors (KNN), then the majority of the KNN is taken to be used as a prediction from the testing data. KNN has several advantages, namely toughness on training data which has a lot of noise and is effective when the training data is large. Whereas.

2.2. Decision Support System

The concept of a Decision Support System (DSS) is expressed by the term Management Decision System. The system is a computer-based system intended to assist decision makers. SPK is an interactive system that helps decision makers through the use of data and decision models to solve semi-structured and unstructured problems.

2.3. Visual Studio Code

Visual Studio Code (VS Code) is a lightweight and powerful text editor made by Microsoft for multiplatform operating systems, meaning that it is also available for Linux, Mac and Windows versions. This text editor directly supports the JavaScript, Typescript, and Node.js programming languages, as well as other programming languages with the help of plug-ins that can be installed via the Visual Studio Code marketplace (such as C++, C#, Python, Go, Java, etc.) [3].

2.4. Website

Website is the Web which can be interpreted as a collection of pages that display information on text data, image data, animation data, sound, video and a combination of all of them, both static and dynamic, which form a series of interrelated buildings, each of which is connected to a network. -page network [4].

2.5. Understanding PHP

PHP stands for Personal Home Page which is the standard language used in the web world. PHP is a programming language in the form of a script that is placed on a web server. PHP can be interpreted as a Hypertext Preprocessor. This is a language that can only run on the server whose results can be displayed on the client. The PHP interpreter in executing PHP code on the server side is called the server side, in contrast to the Java virtual machine that executes programs on the client side [5], [6], [7].

2.6. Sugarcane Plant

Diabetes is a chronic disease because the pancreas cannot produce more. Sugarcane (*Saccharum Officinarum* L) is an annual plantation plant, which has its own characteristics, because there is sugar in the stem. Sugarcane belongs to the grass family (graminae) such as rice, glagah, corn, bamboo and others [1]. Sugar cane which in English is called sugar cane has a high sugar content. Therefore, sugar cane is known for its utilization as a staple for sugar production. Sugarcane is a genus of saccarum which has the highest sucrose content and the lowest fiber content [2], [8], [9].

3. Results And Discussion

The purpose of this research is to classify the quality of the sugarcane produced whether the sugarcane is ripe or immature.

3.1. Calculation of K-NN

The K-Nearest Neighbor (KNN) algorithm is an instance-based-learning group. This algorithm is also a lazy learning

technique. K-NN is done by looking for groups of k objects in the training data that are closest (similar) to objects in the new data or data testing (Wahyudi, 2018).

In K-NN the following formula can be used:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots\dots\dots(1)$$

Information :

x_i = sample data

y_i = test data

i = data variable

d = distance

n = dimensions

The steps of the KNN algorithm are:

- 1) Determining the K parameter (the number of closest neighbors) The K parameter in testing is determined based on the optimum K value during training. The optimum K value is obtained.
- 2) Calculates the square of the Euclidean distance for each object against the given sample data.
- 3) Sort these objects into groups that have the smallest Euclidean distance.
- 4) Gathering category Y (nearest neighbor classification).
- 5) By using the majority category, the calculated query instance value can be predicted.

As for this research, using the K-Nearest Neighbor method in data classification. In this sub-chapter the author uses training data from harvest data in research and development of sugar cane and tobacco with a total of 5 records of data and 1 record of test data. The type of sugarcane used was only one type of sugarcane, namely the Bz134 variety.

Table 1: Train Data

No	Name	Long	Diameter	Section	Classification
1	Bz1341	29	2,6	24	Already Cooked
2	Bz1342	28	2,5	25	Already Cooked
3	Bz1343	28	2,5	26	Already Cooked
4	Bz1344	27	2,5	27	Not yet cooked
5	Bz1345	25	2.0	30	Not yet cooked

Table 2: Test Data

No	Length(cm)	Diameter(cm)	Number of Sections	Classification
1	24	2,5	26	?

The following are the steps for calculating the K-Nearest Neighbor method in data classification.

1. Determined the value of K = 3

2. Calculate the Euclidean distance of the test data.

This process is to calculate the entire value in the sugarcane data to produce the Euclidean distance value.

Table 3: Calculte

No	Name	Euclidean distance (24, 25, 2/26)
1	Bz1341	$\sqrt{((29 - 24)^2 + (2.6 - 2.5)^2 + (24 - 26)^2)} = 5.3860932037981$
2	Bz1342	$\sqrt{((28 - 24)^2 + (2.5 - 2.5)^2 + (25 - 26)^2)} = 4.1231056256177$
3	Bz1343	$\sqrt{((28 - 24)^2 + (2.5 - 2.5)^2 + (26 - 26)^2)} = 4$
4	Bz1344	$\sqrt{((27 - 24)^2 + (2.5 - 2.5)^2 + (27 - 26)^2)} = 3.1622776601684$
5	Bz1345	$\sqrt{((25 - 24)^2 + (2.0 - 2.5)^2 + (30 - 26)^2)} = 4.153311931459$

3. Sort distances in ascending order

At this stage calculations are carried out with the value of length, diameter and number of segments in the cane. Then the calculation is carried out based on the smallest value then the largest value.

Table 4: Sort DDistance

No	Name	Euclidean distance (24, 25, 2/26)	Ascending
1	Bz1341	$\sqrt{((29 - 24)^2 + (2.6 - 2.5)^2 + (24 - 26)^2)} = 5.3860932037981$	5
2	Bz1342	$\sqrt{((28 - 24)^2 + (2.5 - 2.5)^2 + (25 - 26)^2)} = 4.1231056256177$	3

3	Bz1343	$\sqrt{((28 - 24)^2 + (2.5 - 2.5)^2 + (26 - 26)^2)} = 4$	2
4	Bz1344	$\sqrt{((27 - 24)^2 + (2.5 - 2.5)^2 + (27 - 26)^2)} = 3.1622776601684$	1
5	Bz1345	$\sqrt{((25 - 24)^2 + (2.0 - 2.5)^2 + (30 - 26)^2)} = 4.153311931459$	4

4. Determine the shortest distance

In this process, it is done by looking at the most neighbors according to the K value that has been set at the beginning, which is selected by 3 data based on the value of *Euclidean distance*/

Table 5: Determinate

No	Name	Euclidean distance (24, 25, 2/26)	Ascending	Classification
1	Bz1341	$\sqrt{((29 - 24)^2 + (2.6 - 2.5)^2 + (24 - 26)^2)} = 5.3860932037981$	5	Already Cooked
2	Bz1342	$\sqrt{((28 - 24)^2 + (2.5 - 2.5)^2 + (25 - 26)^2)} = 4.1231056256177$	3	Already Cooked
3	Bz1343	$\sqrt{((28 - 24)^2 + (2.5 - 2.5)^2 + (26 - 26)^2)} = 4$	2	Already Cooked
4	Bz1344	$\sqrt{((27 - 24)^2 + (2.5 - 2.5)^2 + (27 - 26)^2)} = 3.1622776601684$	1	Not yet cooked
5	Bz1345	$\sqrt{((25 - 24)^2 + (2.0 - 2.5)^2 + (30 - 26)^2)} = 4.153311931459$	4	not yet cooked

5. Find the most neighboring classes

After selecting a value based on the value of the Euclidean distance, 3 data are obtained with the names namely cane 4, cane 3 and cane 2.

Table 6: Find the most

No	Name	Euclidean distance (24, 25, 2/26)	Classification
1.	Bz1344	$\sqrt{((27 - 24)^2 + (2.5 - 2.5)^2 + (27 - 26)^2)} = 3.1622776601684$	Not yet cooked
2.	Bz1343	$\sqrt{((28 - 24)^2 + (2.5 - 2.5)^2 + (26 - 26)^2)} = 4$	Already Cooked
3.	Bz1342	$\sqrt{((28 - 24)^2 + (2.5 - 2.5)^2 + (25 - 26)^2)} = 4.1231056256177$	Already Cooked

6. Results from X

Judging from the data obtained, there are two data classifications that have good value, then sugarcane with the name X produces a Good classification value. Which means the sugarcane is included in the good sugarcane grouping.

Table 7: Result

No	Name	Long	Diameter	Section	Classification
1.	X	24	2,5	26	Already Cooked

4. Conclusion

Based on the results of making an application to determine the quality of sugarcane grouping, determining whether sugarcane is good or not using the K-Nearest Neighbor method is to calculate the values of sugarcane and look for the closest neighbor classification from sugarcane data that has been previously inputted to give good sugarcane yields or not. In order to classify the best sugarcane data, several supporting values are needed, namely the length of the cane, the diameter of the stem and the number of internodes which are then calculated for the entire value in order to provide results from the classification of sugarcane.

5. Suggestion

The suggestions that the writer can convey are as follows:

1. For now the data that can be managed in the system is for good or bad sugarcane classification and for one type of

sugarcane only. In the future, it is necessary to add other types of sugar cane so that the system can have several choices when carrying out the calculation process.

2. The system that has been designed is still offline, so the system can only run on certain computers. In the future, it is necessary to make an online-based system so that the application can be used anywhere

References

- [1] Nurmuslimah, S. (2020). Aplikasi Metode Fuzzy Mamdani Untuk Pemilihan Tebu Berkualitas Pada Produksi Gula. *Network Engineering Research Operation*, 5(1), 5.
- [2] Hikma, N., Yahya, M., & Zain, S. G. (2022). Pengembangan Aplikasi Taksasi Tebu Berbasis Android Untuk Memprediksi Hasil Panen Tebu Di Kabupaten Takalar.
- [3] Permana, Y., & Romadlon, P. (2019). Perancangan Sistem Informasi Penjualan Perumahan Menggunakan Metode Sdlc Pada Pt. Mandiri Land Prosperous Berbasis Mobile. *Jurnal Teknologi Pelita Bangsa*, 84(10), 1511–1518.
- [4] Destiningrum, M., & Adrian, Q. J. (2017). Sistem Informasi Penjadwalan Dokter Berbasis Web Dengan Menggunakan Framework Codeigniter (Studi Kasus: Rumah Sakit Yukum Medical Centre). *Jurnal Teknoinfo*, 11(2), 30.
- [5] Trimarsiah, Y., & Arafat, M. (2017). Analisis Dan Perancangan Website Sebagai Sarana Informasi Pada Lembaga Bahasa Kewirausahaan Dan Komputer Akmi Baturaja. *Jurnal Ilmiah Matrik*, Vol. 19 No, 1–10.
- [6] Akbar, R., & Dahlan, A. (2013). Pembuatan Sistem Informasi Penjualan Pada Ade Jaya Ponsel Dengan Menggunakan Visual Basic. *Data Manajemen Dan Teknologi Informasi (Dasi)*, 14(4), 39.
- [7] Anwar, S. N., Fatkhul Amin, & Nugroho, S. (2014). Desain Uml Aplikasi Navigasi Layanan Kesehatan Berbasis Android. *Seminar Nasional Sistem Informasi Indonesia*, (September), 250–254.
- [8] Aprianti, & Maliha. (2016). Sistem Informasi Kepadatan Penduduk Kelurahan Atau Desa Studi Kasus Pada Kecamatan Bati-Bati Kabupaten Tanah Laut. *Jurnal Sains Dan Informatika*, 2(2013), 21–28.
- [9] Fridayanthie, E. W., & Mahdiati, T. (2016). Rancang Bangun Sistem Informasi Permintaan Atk Berbasis Intranet (Studi Kasus: Kejaksaan Negeri Rangkasbitung). *Jurnal Khatulistiwa Informatika*, 147(March), 11–40.