JoCoSiR
Journal of Computer Science Research ISSN 2986-2337 (Online)

Application of Data Mining in Classification Fresh Milk Production by Province Using K-Means Algorithm

Afifah Wulandari [a], *M. Safii*[b]

[a],[b]*Information System, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia*

email: [a] *Afifahwulandari.aw@gmail.com,* [b] *m.safii@amiktunasbangsa.ac.id*

ARTICLE INFO

Article history:
Received Dec 28, 2022
Revised Jan 08, 2023
Accepted Jan 22, 2023
Available online Jan 30, 2023

Keywords:
Application
Data Mining
K-Means

IEEE style in citing this article:
A.. Wulandari and M. Safil, " Application of Data Mining in Classification Fresh Milk Production by Province Using K-Means Algorithm, Volume 1 No 1 January 2023, pp. 29-33, 2023.

ABSTRACT

The need for fresh milk is currently experiencing a fairly rapid development as can be seen in terms of the domestic market, here researchers want to increase the productivity and quality of fresh milk production. The data to be used is data from the Central Bureau of Statistics. The method in this study is the K-means clustering algorithm which is grouped into 2 clustering, namely high and low. The results of this study are 1 high-level cluster province, 24 low-level cluster provinces

1. Introduction

Production is all activities that aim to increase or increase the use of an object or all activities aimed at satisfying other people through exchange in covering every human effort and ability to increase benefits in meeting human needs[1]. In production business activities it is a mandatory activity, because with production the company can run its business and further earn profit. The existence of a lot of production and sustainable can increase efforts in developing the business. The existence of a business that continues to increase, one of the indicators is that the business produces in large quantities. This can be seen from a company's microeconomic perspective. However, in essence, from a macroeconomic point of view, they are almost the same when viewed from the aggregate (overall) point of view of the company[2].In the production process certainly has many problems encountered. These problems will later have a negative impact on the company both in terms of product quality and the effectiveness of a process. Therefore the company must analyze the problems that occur by identifying the causes of the problems so that they can prevent these problems from happening in the future[3].

Milk is a food ingredient which is a very good source of nutritional value. Every year the need for milk continues to increase, in line with the increasing population and awareness of the community's nutritional needs which are very important and supported by science and technology. The increased milk production of dairy cows has resulted in not being able to meet the needs of domestic milk consumption, so that milk and processed milk products are imported to meet the nutritional needs of the community[4]The cause of low milk production is closely related to pregnancy failure or miscarriage. This can lead to variations in the age of first childbearing. The non-uniformity of the age of the first child will cause differences in the total milk production produced during each lactation period[5].

Review of previous research (Rosmini, 2018) entitled "Implementation of the K-Means Method in mapping student groups through Lecture Activity Data" concluded that, Cluster A is students who graduate on time while cluster B is students who graduate not on time, and grouping data These students are input for guardian lecturers in guiding and supervising the student learning process so that they can graduate on time.

It is hoped that this research can provide benefits and recommendations for the government to further increase the production of pressed milk in Indonesia so that livestock in Indonesia can meet their individual needs to the fullest**.**

ISSN 2986-2337 (Online)

2. Method

2.1 Data Mining

Data Mining is a process or activity of collecting a large amount of data to be extracted so that it becomes usable information. Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from large databases. The main purpose of data mining is to find, explore, or mine knowledge from the data or information that we have[6]

2.2 Clusters

According to Baskoro cluster or clustering is one of the tools in data mining which aims to group objects into clusters - clusters. Cluster is a data analysis method, which is often included as one of the data mining methods, the purpose of which is to group data with the same characteristics into the same region and data with different characteristics into other regions. Clusters differ from classifications in that clusters do not have a target variable. The purpose of clusters is not to classify, estimate, or predict the value of the target variable[7].

2.3 K-Means

The K-Means algorithm is a group analysis method that leads to the partitioning of N observation objects into groups (clusters) where each object of observation belongs to a group with the closest mean (average).[8].
In general, the K-Means algorithm has steps in grouping:

a. Initialization: determine the desired K centroid value and the desired dissimilarity (distance) metric
b. Selects K data from the X set as the centroid. To determine the centroid, you can use the equation

$$\frac{\boldsymbol{jumlah\ data}}{\boldsymbol{jumlah\ class} + \mathbf{1}}$$

c. Allocating all data to the nearest centroid with a predetermined distance matrix
d. Recalculating the centroid C based on the data that follows each cluster
e. Repeat steps 3 and 4 until convergence is reached

3. Results and Discussion

Based on the description of previous research work steps, several stages will be carried out in the use of the K-Means Clustering method as a solution to solving research cases carried out in accordance with the work steps.

3.1 Data analysis

Data analysis is a method used to answer research problems through management procedures, especially problems related to research. The data obtained for this research is fresh milk production from 2020-2021

Table 1 . BPS Research Data

Province	2020	2021
aceh	9.70	9,11
North Sumatra	3566,16	3624,48
West Sumatra	1024.55	1081.85
Riau	45,37	45,37
Jambi	6,26	6,26
South Sumatra	31,13	32,37
Bengkulu	272.35	263.86
Lampung	2134.84	2168,29
Kep. Bangka Belitung	418.58	426.78
Kep. Riau	12.55	10.45
DKI Jakarta	4292.68	4335.59
West Java	281198.94	283361,18
Central Java	102707,27	102508,64
In Yogyakarta	5385.60	5306.04
East Java	542860.27	556431.78
Banten	121,22	121,22
East Nusa Tenggara	46,87	48,43
West Kalimantan	115.78	116,47
South Kalimantan	111.64	159,19
East Kalimantan	122,21	117.99
North Kalimantan	6,27	8.36
North Sulawesi	13.67	14,19
South Sulawesi	2302,11	2362.75
Southeast Sulawesi	106,63	114.99

a. Determine the value of k for the number of clusters of corn crop production as many as 2 clusters (k-2). The clusters formed are the highest cluster and the lowest cluster

http://doi.org/10.XXXXX/JoCoSiR.v1iss1.pp29-33

ISSN 2986-2337 (Online)

Table 2. The results of calculating the data distance to the cluster center point (Iteration 1)

Province	2020	2021	C1	C2
aceh	9.70	9,11	777362.80	4.47
North Sumatra	3566,16	3624,48	772291.43	5075,86
West Sumatra	1024.55	1081.85	775886,26	1481.15
Riau	45,37	45,37	777311.94	55,31
Jambi	6,26	6,26	777367.25	0.00
South Sumatra	31,13	32,37	777331,19	36.06
Bengkulu	272.35	263.86	776997.04	370.35
Lampung	2134.84	2168,29	774333,27	3034.01
Kep. Bangka Belitung	418.58	426.78	776778.31	588.94
Kep. Riau	12.55	10.45	777359.85	7.56
DKI Jakarta	4292.68	4335.59	771275.08	6092,33
West Java	281198.94	283361,18	378198,63	399198,36
Central Java	102707,27	102508,64	632282,28	145100.78
In Yogyakarta	5385.60	5306.04	769817.25	7551.49
East Java	542860.27	556431.78	0.00	777367.25
Banten	121,22	121,22	777204,68	162.58
East Nusa Tenggara	46,87	48,43	777308,70	58,54
West Kalimantan	115.78	116,47	777211.88	155,37
South Kalimantan	111.64	159,19	777184,19	185.72
East Kalimantan	122,21	117.99	777206,30	161.02
North Kalimantan	6,27	8.36	777365,74	2,10
North Sulawesi	13.67	14,19	777356.40	10.85
South Sulawesi	2302,11	2362.75	774077,27	3289.98
Southeast Sulawesi	106,63	114.99	777219.33	147.97

b. Determine the initial centroid (cluster center) value that has been determined based on the value of the data variable in the cluster as much as previously determined. The highest cluster is obtained from the highest value on, the low cluster is taken from the average value and the lowest cluster

Table 3.Centroid initial data (Iteration 1)

MAX	556431.78
min	6,26

c. Calculating the distance of each fresh milk production data to the cluster center. After the initial cluster center value data is determined, the next step is to calculate the distance of each data to the cluster center using the formula whose calculations can be seen as follows: Calculation of the distance to the fresh milk production data with the centroid in the first cluster

Table 4. Cluster Center and Center Point (Centroid)

C1	1787,93	1816,795
C2	42878,945	43592,82045

d. Determine the cluster position of each corn production data based on the minimum data distance to the cluster center. Data that has the smallest distance from the centroid will be a member of that group. The following is a table of data positions with each cluster in the 1st iteration using the help of a sign (1) which says data is a member of the cluster

Table 5. Position of data grouping results for each cluster (Iteration 1)

Province	C1	C2
aceh		1
North Sumatra		1
West Sumatra		1
Riau		1
Jambi		1
South Sumatra		1
Bengkulu		1
Lampung		1
Kep. Bangka Belitung		1
Kep. Riau		1
DKI Jakarta		1
West Java	1	
Central Java	1	
In Yogyakarta		1

http://doi.org/10.XXXXX/JoCoSiR.v1iss1.pp 29-33

ISSN 2986-2337 (Online)

East Java	1	
Banten		1
East Nusa Tenggara		1
West Kalimantan		1
South Kalimantan		1
East Kalimantan		1
North Kalimantan		1
North Sulawesi		1
South Sulawesi		1
Southeast Sulawesi		1

e. Then do steps 3 to 4 again. If the centroid value from the iteration results with the previous centroid value is the same or the centroid value is optimal and the position of the corn plant production data cluster does not change anymore, the iteration process stops. However, if the centroid values are not the same or not optimal and the position of the corn crop production data is still changing, then the iteration process continues in the next iteration. Iteration 2

Table 6 . Result of calculation of data distance to cluster center point (Iteration 2)

Province	2020	2021	C1	C2	C1	C2
Aceh	9.70	9,11	2535,710353	61133.56		1
North Sumatra	3566,16	3624,48	2535,710353	56062,14		1
West Sumatra	1024.55	1081.85	1059,666541	59657,13		1
Riau	45,37	45,37	2484,846451	61082.69		1
Jambi	6,26	6,26	2540,154518	61138.00		1
South Sumatra	31,13	32,37	2504,100401	61101.95		1
Bengkulu	272.35	263.86	2169.928536	60767,76		1
Lampung	2134.84	2168,29	493.8575535	58103.99		1
Kep. Bangka Belitung	418.58	426.78	1951,220419	60549.07		1
Kep. Riau	12.55	10.45	2532,756681	61130,60		1
Dki Jakarta	4292.68	4335.59	3552,196618	55045.71		1
West Java	281198.94	283361,18	396658,2323	338061.07	1	
Central Java	102707,27	102508,64	142560,7268	83967,27	1	
In Yogyakarta	5385.60	5306.04	5011,792105	53587,58		1
East Java	542860.27	556431.78	774827,1154	716229,80	1	
Banten	121,22	121,22	2377,582134	60975.43		1
East Nusa Tenggara	46,87	48,43	2481,613325	61079.46		1
West Kalimantan	115.78	116,47	2384,783162	60982,63		1
South Kalimantan	111.64	159,19	2357,456786	60955.09		1
East Kalimantan	122,21	117.99	2379,193466	60977.04		1
North Kalimantan	6,27	8.36	2538,651119	61136.50		1
North Sulawesi	13.67	14,19	2529,304911	61127,15		1
South Sulawesi	2302,11	2362.75	749.9652888	57848.06		1
Southeast Sulawesi	106,63	114.99	2392,260426	60990.10		1

4. Conclusions

The conclusion that can be drawn from data mining techniques in clustering fresh milk production by province with the K-Means algorithm is that this K-Means Clustering Algorithm can help researchers classify fresh milk production in each province in Indonesia. From fresh milk production data in Indonesia It is known that the province with the highest cluster is East Java Province.

5. References

[1] U. Duwila, "The Influence of Rice Production on the Level of Welfare of the People of Waeapo District, Buru Regency,"Econ's dream. J. Ekon., vol. IX, no. 2, pp. 149–158, 2015.

[2] A. Muhtarom, UI Lamongan, JN Investment, M. Di, and J. Timur, "Development of Total Production Value and Amount," vol. I, no. 2, pp. 83–93, 2016.

[3] DS Bowo, "Analysis of Production Process Improvements at PT Sumber Teknik Sentosa,"J. business management, vol. 8, no. 01, pp. 19–28, 2018.

[4] TA Prabowo, S. Indarjulianto, A. Pertiwiningrum, and C. Sugiyanto, “Reproduction Performance and Milk Production of Dairy Cows in Tegalombo District, Pacitan Regency, East Java Province. ,” vols. 10, no. 1, pp. 29–36, 2021.

[5] K. Murdani, M. Jakfar, and UJ Ghafur, "DAIRY CATTLE PRODUCTIVITY BASED ON LENGTH OF CARE AND TOTAL MILK PRODUCTION," vol. 12, no. September, 2022.

[6] N. Erlangga, "APPLEMENT OF DATA MINING IN GROUPING CORN PRODUCTION BY PROVINCE USING THE K-MEANS ALGORITHM," vol. 3, pp. 702–709, 2019, doi: 10.30865/komik.v3i1.1681.

[7] M. Production and T. Medicine, “IMPLEMENTATION OF BUSINESS INTELLIGENCE SYSTEM AND K-

http://doi.org/10.XXXXX/JoCoSiR.v1iss1.pp29-33

ISSN 2986-2337 (Online)

MEANS CLUSTERING TO," vol. 32, no. 2, pp. 204–219, 2022.

[8] S. Natalia, B. Sembiring, H. Winata, and S. Kusnasari, "Classifying Student Achievement Using the K-Means Algorithm," vol. 1, pp. 31–40, 2022.