Jurnal Mantik, 5 (4) (2022) 2452-2461

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik/index

Application of K-Means Algorithm Data Mining in Goat Meat Production Data Grouping in Indonesia

Mochamad Wahyudi[1], Solikhun[2], Lise Pujiastuti[3]

[1]Universitas Bina Sarana Informatika, Jl. Kramat Raya No. 98 Senen, Jakarta Pusat, DKI Jakarta, Indonesia
[2]AMIK Tunas Bangsa, Jl. Jendral Sudirman Blok A No. 1/2/3 Pematangsiantar, Sumatera Utara, Indonesia
[3]STMIK Antar Bangsa, Kawasan Bisnis CBD Ciledug, Jl. HOS Cokroaminoto No. 29-35 Tangerang, Banten, Indoensia

Email: wahyudi@bsi.ac.id[1], solikhun@amiktunasbangsa.ac.id[2], lise.pujiastuti@antarbangsa.ac.id[3]

ARTICLEINFO

Article history:

Received: Dec 17, 2021
Revised: Jan 18, 2022
Accepted: Feb 08, 2022

Keywords:

Data mining,
K-means Clustering,
Goat Meat Production

ABSTRACT

Data mining is the process of mining data from big data to get important information. The data mining process requires the use of artificial intelligence technology. The production of goat meat is very much needed in the fulfillment of protein ingredients for the people of Indonesia. It is necessary to make a grouping of goat meat production to see the condition of the map of the strength of meat production in Indonesia, so that the government can take appropriate steps to develop goat meat production in Indonesia. This study uses data mining techniques using the k-means clustering method to classify goat meat production in Indonesia. The results of this study are data on mutton product clustering, namely 2 nodes in the high group, the low group having 22 nodes, and the medium group having 10 nodes.

1. Introduction

Data mining is a process or activity to collect a large amount of data for extraction 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 [1].

Clustering is a data mining technique used to analyze data that has variations and the number of lots. Clustering is the process of grouping data into a cluster, so that it contains data that is as similar as possible and different from other cluster objects [2]. Clustering exists in everyday life, because it cannot be separated from a number of data that produce information to meet the needs of life. One of the most important tools in relation to data is to classify or classify data into a set of categories or clusters [3].

The clustering problem has been studied extensively in computer science. They play a central role in many fields, including data science and machine learning, and their studies have led to the development and refinement of several key techniques in algorithmic and theoretical computer science [4].

The production of goat meat in Indonesia needs to be considered considering that goat meat is one of the foodstuffs that produces protein to fulfill the nutrition of the Indonesian people. The government needs to

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

conduct a study to determine the mapping of the state of kabob meat production in Indonesia in order to continue to increase goat meat production in Indonesia. The purpose of this study was to determine the classification of meat production data in Indonesia. The results of this study serve as input for the government to determine the right program to increase goat meat production in Indonesia. The method used in this study is the k-means clustering method by dividing into 3 groups, namely low, medium and high groups.

2. Literature Review

2.1 Data Mining

Data mining is often called knowledge discovery in database (KDD), activities that include collecting, using historical data to find regularities, patterns or relationships in large data sets [5]. Data mining is a term used to describe the discovery of knowledge in databases. 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[6].

2.2 Cluster Analysis

Cluster analysis is a multivariate technique where the analysis process uses a clustering algorithm rather than being done by humans, with the main purpose of this analysis being to sort data based on their characteristics so that data that have the closest resemblance to other data are grouped into one cluster. In other words, this analysis is useful for directing the findings of groups that were previously unknown to the data [7].

2.3 K-Means

K-Means technique is often used in Data Mining, because of its ease in handling and classifying large amounts of data through clustering or grouping. Clustering means grouping things that are similar or have the same features, and so is the purpose of k-means clustering. K-means clustering is an unattended machine learning algorithm for clustering 'n' observations into clusters 'k' where k is a predefined or user-defined constant. The main idea is to define k centroids, one for each cluster. The K-Means algorithm involves: 1. Selecting the number of "k" clusters. 2. Randomly assign each point to a cluster. 3. Until the clusters stop changing, repeat the following steps: For each cluster, calculate the centroid of the cluster by taking the vector mean of the points in the cluster. Assign each data point to the cluster whose centroid is closest. Two things are very important in K-means, the first is to scale the variables before grouping the data, and the second is to look at the scatter plot or data table to estimate the number of cluster centers to be assigned to the k parameter in the model [8].

3. Results and Analysis

3.1 Data analysis

Data analysis is a method used to answer research problems through its management procedures, especially problems related to research. The data obtained for this research is goat meat production data in 2013-2020.

Table 1
BPS/Research Data

province	2013	2014	2015	2016	2017	2018	2019	2020
Aceh	2229	3012	2604	2786	2710	2841	2209	2276
North Sumatra	3470	3538	3546	3959	3981	1213	1118	1187
West Sumatra	646	669	686	692	751	825	718	733
Riau	550	620	648	652	730	827	685	705
Jambi	721	689	658	760	802	1294	1649	1141

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

province	2013	2014	2015	2016	2017	2018	2019	2020
South Sumatra	2222	1532	1300	1395	1868	1999	1713	1697
Bengkulu	247	436	547	258	95	100	95	114
Lampung	2466	2023	1807	2108	2116	1822	2158	1921
Kep. Bangka Belitung	101	96	94	56	77	90	76	77
Kep. Riau	90	281	329	344	348	300	466	472
DKI Jakarta	1263	1133	870	1110	1208	1300	842	842
West Java	7160	7883	8476	7384	9397	10137	7280	4454
Central Java	10211	11174	11051	11669	11857	11820	12549	12177
In Yogyakarta	1490	1483	1598	1653	2159	1602	1920	1769
East Java	15499	16622	16465	17950	18681	21530	25361	25595
Banten	3246	2574	3499	2298	2545	3564	4695	4331
Bali	1679	1728	2599	3403	2912	858	901	991
West Nusa Tenggara	391	213	184	222	276	270	305	323
East Nusa Tenggara	3630	3727	1733	1763	2018	2107	2433	2921
West Kalimantan	515	276	264	358	464	411	417	421
Central Kalimantan	339	315	344	205	200	224	215	222
South Kalimantan	663	654	565	551	441	361	318	271
East Kalimantan	409	533	441	499	585	623	556	573
North Kalimantan	0	35	43	42	50	51	88	101
North Sulawesi	393	395	397	243	187	187	188	131
Central Sulawesi	2460	1510	1608	2861	1354	1117	1396	1409
South Sulawesi	1591	955	971	1039	1023	1125	1109	1143
Southeast Sulawesi	265	124	317	318	366	356	313	314
Gorontalo	178	68	158	198	266	317	252	229
West Sulawesi	273	272	503	211	211	213	168	184
Maluku	450	203	170	342	218	196	237	235
North Maluku	35	28	99	122	114	130	82	78
West Papua	155	156	157	158	63	74	71	74
Papua	132	185	216	234	281	271	268	293

3.2 Defines a manually defined or random Centroid value extracted from the data.

Table 2

Centroid Value

Centroid	2013	2014	2015	2016	2017	2018	2019	2020
Centroid 1 (Medium)	2543,556	2347,444	2254,889	2469,556	2407,000	1902.556	2060,333	2055,778
Centroid 2 (Low)	427,591	378,909	393,682	391,545	398.000	433,864	414,455	394,364
Centroid 3 (High)	10956667	11893,000	11997.333	12343333	13311,667	14495,667	15063,333	14208.667

3.3 Calculating distance from Centroid

To calculate the distance between the centroid point with the point of each object using Euclidion Distance.

$$D_{(i,f)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \cdots + (X_{ki} - X_{kj})^2}$$

Then the calculation for the distance from the 1st Centroid is as follows:

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

$$Dx1,c1=\sqrt{(2229-2543.556)^2+(3012-2347.444)^2+(2604-2254.889)^2+(2786-2469.556)^2+(2710-2407.000)^2+(2841-1902.556)^2+(2209-2060.333)^2+(2276-2055.778)^2}=1343.76$$

$$Dx2,c1=\sqrt{(3470-2543.556)^2+(3538-2347.444)^2+(3546-2254.889)^2+(3959-2469.556)^2+(3981-2407.000)^2+(1213-1902.556)^2+(1118-2060.333)^2+(1187-2055.778)^2}=3279.77$$

And so on up to Dx34,c1. Furthermore, the calculation for the distance from the 2nd Centroid is as follows:

$$Dx1,c2=\sqrt{(2229-427.591)^2+(3012-378.909)^2+(2604-393.682)^2+(2786-391.545)^2+(2710-398.000)^2+(2841-433.864)^2+(2209-414.455)^2+(2276-394.364)^2}=6220.75$$

$$Dx2,c2=\sqrt{(3470-427.591)^2+(3538-378.909)^2+(3546-393.682)^2+(3959-391.545)^2+(3981-398.000)^2+(1213-433.864)^2+(1118-414.455)^2+(1187-394.364)^2}=7514.53$$

And so on up to Dx34,c2. Furthermore, the calculation for the distance from the 3rd Centroid is as follows:

$$Dx1,c3=\sqrt{(2229-10956.667)^2+(3012-11893.000)^2+(2604-11997.333)^2+(2786-12334.333)^2+(2710-13311.667)^2+(2841-14495.667)^2+(2209-15063.333)^2+(2276-14208.667)^2}=29837.87$$

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

$$\sqrt{((3470-10956.667)^2+(3538-11893.000)^2)+(3546-11997.333)^2+(3959-12334.333)^2+(3981-13311.667)^2+(1213-14495.667)^2+(1118-15063.333)^2+(1187-14208.667)^2} = 29915.78$$

Dx2,c3=

And so on up to Dx34, c3. So that we get the distance table from the Centroid and look for the minimum value of the three Centroids. The distance table from the Centroid is as follows:

Table 3
1st Iteration Centroid Distance

C1	C2	C3	Centroid Distance
2797.31	5868.10	25491.68	2797.31
4785.06	7251.63	25406.82	4785.06
2940.24	540.50	30739.92	540.50
3051.88	436.36	30843.74	436.36
2392.83	1500.50	30010.48	1500.50
0.00	3459.36	27944.78	0.00
4339.95	1020.69	32116.23	1020.69
1186.20	4367.06	26992.51	1186.20
4686.60	1273.21	32529.97	1273.21
4050.99	616.95	31819.53	616.95
1938.78	1627.05	29759.73	1627.05
17708.40	20976.29	11892.11	11892.11
27944.78	31266.06	0.00	0.00
991.93	3386.48	27917.94	991.93
52170.77	55443.70	24946.77	24946.77
5150.07	8279.31	23354.65	5150.07
3098.76	4679.21	27618.22	3098.76
4136.18	768.83	31994.63	768.83
3028.88	6093.88	25746.39	3028.88
3800.11	480.11	31661.20	480.11
4208.24	833.77	32044.34	833.77
3626.68	578.33	31438.59	578.33
3459.36	0.00	31266.06	0.00
4782.95	1355.64	32614.81	1355.64
4203.29	875.25	32032.13	875.25
1877.95	3787.08	28025.50	1877.95
1765.13	1795.07	29626.29	1765.13
4094.71	693.93	31921.88	693.93
4329.35	920.01	32169.66	920.01
4246.75	885.01	32056.77	885.01
4193.00	861.22	32048.67	861.22
4686.03	1260.39	32518.15	1260.39
4614.81	1209.00	32449.22	1209.00
4274.26	834.47	32092.62	834.47
2797.31	5868.10	25491.68	2797.31
4785.06	7251.63	25406.82	4785.06

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

C1	C2	C3	Centroid Distance
2940.24	540.50	30739.92	540.50

3.4 Determining Clusters or Grouping

In determining the Cluster by looking for the Cluster value based on the minimum value of the Cluster value and placing it in the Cluster that corresponds to the minimum value in Iteration 1. Here is the Cluster table in Iteration 1 as follows:

Table 4.
1st Iteration Cluster

C1	C2	C3
1		
1		
	1	
	1	
	1	
1		
	1	
1		
	1	
	1	
	1	
		1
		1
1		
		1
1		
1		
	1	
1		
	1	
	1	
	1	
	1	
	1	
	1	
1		
	1	
	1	
	1	
	1	
	1	
	1	
	1	
	1	

Furthermore, in the K-Means method, the calculation stops if the cluster in the resulting iteration is the same as in the previous iteration. Then look for the cluster in the next iteration until the iteration value is the same. To find the next Centroid value, use the new Centroid in the 1st Iteration by adding up the values as

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

listed in the Cluster in the table above. The new Centroid to find the next Cluster is to add up the selected values in the Cluster and then distribute it as many as the number of values.

The data for the new Centroid 1st iteration are as follows:

Table 5
New Centroid 1st Iteration

Centroid	2013	2014	2015	2016	2017	2018	2019	2020
Cluster 1 (Medium)	2543,556	2347,444	2254,889	2469,556	2407,000	1902.556	2060,333	2055,778
Cluster 2 (Low)	427,591	378,909	393,682	391,545	398.000	433,864	414,455	394,364
Cluster 3 (High)	10956667	11893,000	11997.333	12343333	13311,667	14495,667	15063,333	14208.667

By using the same steps as before to determine the Distance from the Centroid by using the new Centroid of the 1st Iteration, the following is the result of the Distance from the Centroid:

Table 6
2nd Iteration Centroid Distance

C1	C2	C3	Centroid Distance
1343.76	6220.75	29837.87	1343.76
3279.77	7514.53	29915.78	3279.77
4415.21	889.94	35040.23	889.94
4530.92	800.06	35140.86	800.06
3936.64	1836.43	34268.47	1836.43
1839.57	3790.97	32243.51	1839.57
5741.36	679.07	36435.34	679.07
749.32	4695.83	31327.76	749.32
6172.29	908.65	36830.22	908.65
5511.36	398.88	36113.61	398.88
3411.64	1949.12	34074.72	1949.12
16262.55	21321.45	16045.46	16045.46
26434.99	31623.39	4704.89	4704.89
1791.45	3743.17	32223.35	1791.45
50786.46	55808.69	20439.26	20439.26
4125.48	8617.83	27563.40	4125.48
2438.20	4940.76	32045.55	2438.20
5640.88	400.48	36293.77	400.48
2221.65	6404.84	30090.46	2221.65
5310.49	204.37	35960.28	204.37
5676.88	449.38	36351.09	449.38
5040.19	466.42	35762.36	466.42
4943.72	400.70	35563.18	400.70
6268.77	1003.29	36911.72	1003.29
5650.50	499.64	36345.76	499.64
1966.10	4044.30	32408.09	1966.10
3286.24	2085.44	33937.31	2085.44
5582.53	355.73	36218.75	355.73
5835.55	580.33	36462.33	580.33
5691.62	518.28	36365.75	518.28
5674.39	479.38	36355.96	479.38
6169.02	905.74	36816.29	905.74
6080.70	834.79	36755.41	834.79
5757.45	503.29	36388.95	503.29

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

C1	C2	C3	Centroid Distance
1343.76	6220.75	29837.87	1343.76
3279.77	7514.53	29915.78	3279.77
4415.21	889.94	35040.23	889.94

From the Centroid Distance table above, the Cluster or 2nd Iteration grouping is as follows:

Table 7
2nd Iteration Cluster

C1	C2	C3
1		
1		
	1	
	1	
	1	
1		
	1	
1		
	1	
	1	
	1	
		1
		1
1		
		1
1		
1		
	1	
1		
	1	
	1	
	1	
	1	
	1	
	1	
1		
	1	
	1	
	1	
	1	
	1	
	1	
	1	

Then proceed to the next iteration until the cluster value is the same or does not change in the next iteration cluster, the calculation is stopped.

3.5 RapidMiner Display

To get the grouping results at the next stage, click the blue arrow in the upper middle position on the toolbar. This stage will display the final results and the final step in using rapidminer tools. Can be seen in Figure 1:

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

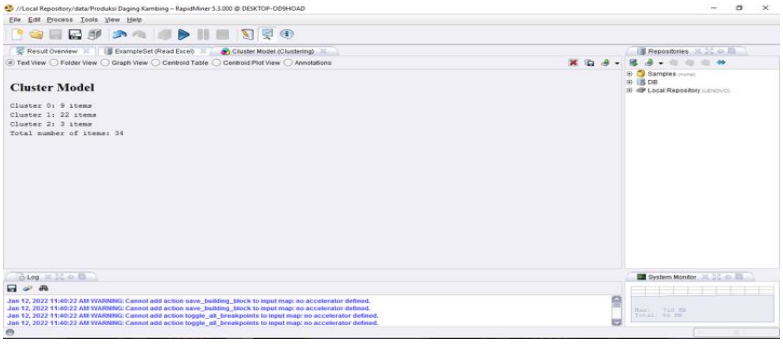

Fig 1. Model Cluster Value

Description :

a. The number of Cluster 1 (Medium) is 9 Items
b. The number of Cluster 2 (Low) is 22 Items
c. The number of Cluster 3 (High) is 3 Items

So that it can be seen the results of the rapidminer grouping can be seen in the following picture:

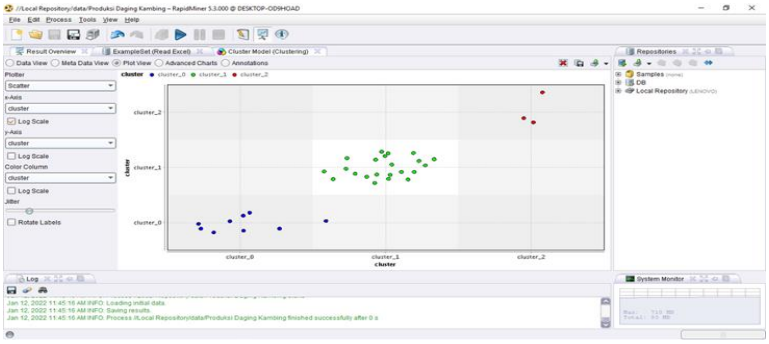

Fig 2. Grouping Results

Based on the figure, it can be seen that the high group has 3 nodes, the low group has 22 nodes, and the medium group has 9 nodes.

4. Conclusion

The conclusions that can be drawn on data mining techniques in clustering goat meat production in Indonesia with the k-means algorithm are as follows:

a. This k-means clustering algorithm can help researchers classify goat meat production in Indonesia. Grouping results. From the goat meat production data in Indonesia, it can be seen that there are 3 provinces with high-level clusters, namely: West Java, Central Java and East Java, 9 provinces with moderate-level clusters, namely Aceh, North Sumatra, South Sumatra, Lampung, DI Yogyakarta. , Banten, Bali, NTT and Central Sulawesi: and 22 provinces with low-level clusters, namely other than provinces with high-level clusters and provinces with moderate-level clusters.
b. This rapidminer application can help researchers classify goat meat production in Indonesia. The manual calculation of the k-means clustering algorithm in classifying goat meat production and its application in Rapidminer showed the same results.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

References

[1] N. Erlangga, S. Solikhun, and I. Irawan, "Penerapan Data Mining Dalam Mengelompokan Produksi Jagung Menurut Provinsi Menggunakan Algoritma K-Means," KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer), vol. 3, no. 1, pp. 702–709, 2019.

[2] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster," IOP Conf. Ser. Mater. Sci. Eng., vol. 336, no. 1, 2018.

[3] S. D. Al-Shobaki, "Jordan journal of mechanical and industrial engineering.," Jordan J. Mech. Ind. Eng., vol. 5, no. 3, pp. 267–272, 2007.

[4] S. Ahmadian, A. Norouzi-Fard, O. Svensson, and J. Ward, "Better guarantees for k-Means and euclidean k-Median by primal-dual algorithms," SIAM J. Comput., vol. 49, no. 4, pp. 97–156, 2020.

[5] S. Al Syahdan and A. Sindar, "Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota," J. Nas. Komputasi dan Teknol. Inf., vol. 1, no. 2, 2018.

[6] L. Maulida, "Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means," JISKA (Jurnal Inform. Sunan Kalijaga), vol. 2, no. 3, p. 167, 2018.

[7] S. S. Nagari and L. Inayati, "Implementation of Clustering Using K-Means Method To Determine Nutritional Status," J. Biometrika dan Kependud., vol. 9, no. 1, p. 62, 2020.

[8] J. E. Ricardo, J. J. D. Menéndez, I. F. B. Arias, J. M. M. Bermúdez, and N. M. Lemus, "Neutrosophic K-means for the analysis of earthquake data in Ecuador," Neutrosophic Sets Syst., vol. 44, pp. 255–262, 2021.