Computer Engineering and Applications Vol. 13, No. 2, June 2024

Application of Machine Learning in Clustering Maize Producing Regions in Indonesia

Eliyani, Saruni Dwiasnati, Sutan Mohammad Arif , Reza Avrizal, Nona Fatimah*

[1,2,5]*Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Mercu Buana, Indonesia*
[3,4]*Fakultas Teknik dan Ilmu Komputer, Program Studi Teknik Informatika, Universitas Indraprasta PGRI, Indonesia*
*eliyani@mercubuana.ac.id, *saruni.dwiasnati@mercubuana.ac.id , cutans.muhars@gmail.com, avrizale.pln@gmail.com, nonaalbinsaid@gmail.com*

ABSTRACT

Maize is considered an important commodity with promising market prospects. Given the importance of maize, there is a need to increase maize production to meet people's needs and maintain price stability. This study aims to group maize production in Indonesia by region, with the hope of finding areas that have the potential to become maize production centers to reduce dependence on imports. The data used in this research was obtained from the Central Statistics Agency, covering information from 34 provinces during the 2017-2021 period. This analysis uses the K-Means method with the Python programming language. The number of groups is determined using the Elbow Method. The results of this research show that there are three categories of maize production regions: regions with low maize production (below average), regions with medium maize production, and regions with high maize production. A total of 25 provinces are in the low production category, eight provinces are in the medium category, and only East Java is in the high production category.

Keywords: Maize; Clustering; K-means; Elbow

1. INTRODUCTION

The right to obtain food is one of the human rights in accordance with the mandate of the 1945 Constitution. Economic stability and national stability can be influenced by adequate food availability. Efforts are made to increase food security in Indonesia primarily from increasing domestic production.

Regionalization of food crop production centers is one approach to securing national food availability. One method for zoning using data mining techniques is clustering. Clustering is the process of dividing a group of data into a number of smaller groups so that group members who are similar will become one cluster and those who are not similar will become another cluster. Similarity indicators are usually calculated using distance, for example the Euclidian method that is often used. The distance between each member in one cluster is as close as possible, while the distance between clusters is as far as possible. This similarity is determined by one or more parameters. Apart from regionalizing food production, clustering is also used for various purposes, for example in the fields of health[1], telecommunications[2], crime[3], geochemistry[4], and other fields.

K-Means is a non-hierarchical clustering method that is unsupervised (no training process and output targets are required) which groups data into one or more groups using a centroid-based partitioning method. Various studies using the K-Means clustering algorithm relating to food crops, especially corn, have been carried out. The K-Means clustering method is applied in this research to produce groups (clusters) of data that can describe patterns of similarity in the characteristics of the

determining qualitative assessment attribute data[5]. The K-Means algorithm is an iterative clustering algorithm which partitions the data set into a number of k clusters that have been determined at the beginning[6].

Tendean explain about clustered 34 provinces in Indonesia based on the production of five food crops, namely maize, peanuts, soybeans, rice and cassava using the K-Means Clustering algorithm and utilizing rapid miner software[7]. In this research, three clusters were obtained. Of the 34 provinces, 27 provinces were in cluster 1 (C1), namely low food production, 4 provinces in cluster 2 (C2), namely medium food production, and 3 provinces in cluster 3 (C3), namely high food production. The three provinces that are classified as high food producers include: West Java, Central Java and East Java.

Based on production data from 2010 to 2015, using the K-Means algorithm and distance calculations using the Euclidian method, utilizing RapidMiner software,[8]found two clusters of maize producing provinces, namely high production with two provinces namely Central Java and East Java and 32 other provinces classified as low production.

Flomina Gutilized the K-Means algorithm to group corn farmer groups in Rangkat District, Tanah Datar Regency, West Sumatra, and obtained three farmer group clusters that took part in Hybrid corn development activities, where there were seven farmer groups that fell into the high production cluster category generally using NK 212 variety[9]. The NK 212 variety is also the best-selling seed based on UD sales data. Tiara Bersaudara from January to April 2019 which was clustered using the K-Means Clustering algorithm using RapidMiner Studio software along with three other seed varieties, namely NK 7328, Pioneer 32, and NK 617232[10]. Using the K-Means Clustering Algorithm,[11] also grouped corn seeds from 142 corn seeds and obtained three clusters as low quality seeds, quality seeds, and very high quality seeds with the attributes of uniform growth, high productivity and climate change resistance.

To detect leaf diseases in corn plants, segmentation between blight and leaf spot diseases has been carried out using the K-Means algorithm using the Euclidian method to calculate the closest distance, and obtained an accuracy of 90% using image data[12].

2. MATERIAL AND METHODS

This research uses the Clustering method with the K-Means algorithm to cluster which provinces are included in the distribution area for maize producers using various maize production attributes. The tools used to process data in this research use Google Colabs.

2.1 RESEARCH STAGES

This research aims to identify the distribution areas producing maize in Indonesia using the Machine Learning method. The research stages can be seen in Figure 1.

ISSN: 2252-4274 (Print) 31
ISSN: 2252-5459 (Online)

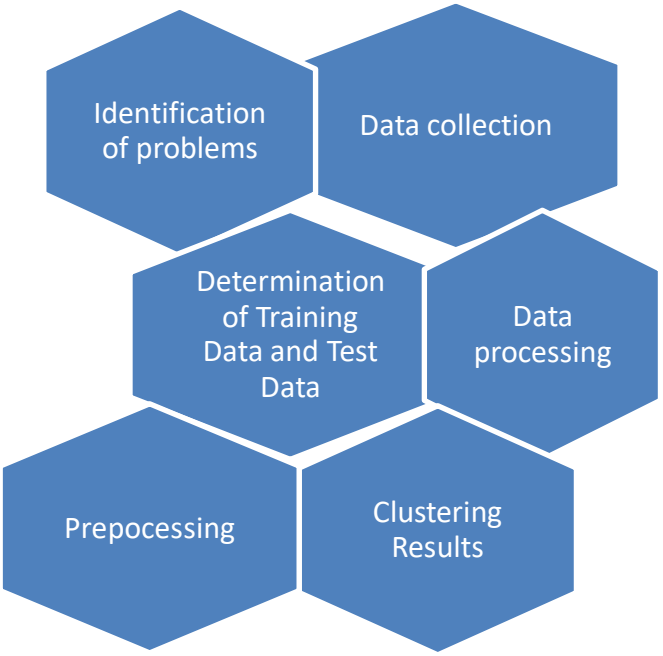

Figure 1. Research Stages

The research stages applied in this research are as follows:

a. Identification of problems

The initial stage of research is to identify the problem that you want to address in order to find the best solution using methods that are appropriate to the data that can be processed to produce good values. Identification of this problem is carried out to be able to consider the difficulty and ease of obtaining data sets from the research being carried out. Apart from that, we also look for data sets that can be used from various official websites related to the objects raised in this research.

b. Data collection

This stage collects data in the form of literature related to data on the distribution of maize producers obtained from the official website of the Ministry of Agriculture throughout Indonesia to find out which areas produce quite a lot of miaze plants each year.

c. Determination of Training Data and Test Data

Determination of training data is based on past data that is available to be used as training material for the algorithm used. Training data obtained from technical instructions for corn producing distribution areas for land evaluation for agricultural commodities prepared by the Center for Research and Development of Agricultural Land Resources.

d. Data processing

This stage processes the distribution data on the website into data that is in accordance with the algorithm used.

e. Prepocessing

In preprocessing there are two stages, namely as follows:

1. Data Cleaning is used to remove unnecessary data such as dealing with missing values, noise data and dealing with inconsistent and relevant data.
2. Data Integration is carried out on attributes that identify unique entities.

f. Clustering Results

The next stage is the results obtained from all the processes above to obtain a clustering of the distribution areas producing maize in Indonesia, the results are in the form of an image.

2.2 RESEARCH METHODS

The data set used in this research comes from the Central Bureau of Statistics, namely annual data on maize production (in tons of dry shelled), corn harvested area (hectares), corn productivity (quintals per hectare), and corn planted area (hectares) from 34 provinces in Indonesia from 2017 to 2021. The top five rows of data from the dataset used are presented in Figure 2.

	no	Provinsi	PJ_2017	PJ_2018	PJ_2019	PJ_2020	PJ_2021	PNJ_2017	PNJ_2018	PNJ_2019	...	PJG_2017	PJG_2018	P
0	1	Aceh	387470	259318	242443	369579	308790	81552	46013	42648	...	47.51	45.22	
1	2	Sumatera Utara	1741258	1227614	1298165	1494380	1452531	281423	211878	217985	...	61.87	61.63	
2	3	Sumatera Barat	985847	662295	538410	687592	745038	142334	102641	82484	...	69.26	70.02	
3	4	Riau	30765	24374	70954	35414	19484	12231	11207	15509	...	25.15	24.88	
4	5	Jambi	98680	69510	58918	60085	41641	15508	9914	9749	...	63.63	60.77	

Figure 2. Data used

Of the thirty-four provinces, only DKI Jakarta does not have a dataset or no data on production, maize harvested area, maize planted area and maize productivity. DKI Jakarta Province does not produce dry shelled corn.

Except for the province name data with the object data type, the other data types are integers, as shown in Figure 3. With this data type, the data can be processed directly.

Data is processed using the Python programming language with the K-Means Clustering algorithm. Determining the number of clusters is approached using the Elbow method.

The aim of this data analysis is to obtain a clustering of corn production areas in Indonesia based on production data for the last five years using the K-Means Clustering algorithm.

```
In [255]: df.dtypes

Out[255]: no           int64
          Provinsi    object
          PJ_2017      int64
          PJ_2018      int64
          PJ_2019      int64
          PJ_2020      int64
          PJ_2021      int64
          PNJ_2017     int64
          PNJ_2018     int64
          PNJ_2019     int64
          PNJ_2020     int64
          PNJ_2021     int64
          PJG_2017   float64
          PJG_2018   float64
          PJG_2019   float64
          PJG_2020   float64
          PJG_2021   float64
          LTJ_2017     int64
          LTJ_2018     int64
          LTJ_2019     int64
          LTJ_2020     int64
          LTJ_2021     int64
          dtype: object
```

Figure 2. Data type

ISSN: 2252-4274 (Print) 33
ISSN: 2252-5459 (Online)

3. RESULT AND DISCUSSION

There was no missing data found in the dataset as presented in Figure 4 so that segmentation could be carried out directly using the K-Means algorithm

Figure 4. The dataset does not have missing data

The number of clusters is determined using the Elbow Method, where the number of clusters to be formed depends on the number of breaks in the curve that forms the elbow. In Figure 5, you can see that the line that is closest to the shape of an elbow is at k=3, which means the best number of clusters is 3 clusters.

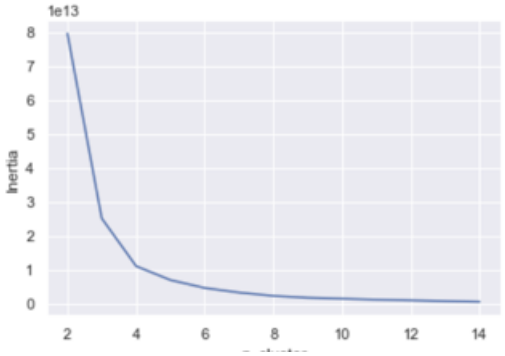

Figure 5. Elbow method curve

These three clusters can be categorized as low maize producing areas (below average), medium maize producing areas and high corn producing areas.

Provinces that are classified in the low cluster along with their average production value during the period 2017 – 2021 are presented in Table 1. Based on Table 1. There are 25 provinces that are included in the low cluster. The range of dry shelled maize production per year from this low cluster is between 0 to 747,461 tons of dry shelled maize.

Table 1. Low cluster

No	Province	PJ(ton PK)
1	Aceh	313,520
2	Sumatera Barat	723,836
3	Riau	36,198
4	Jambi	65,767
5	Sumatera Selatan	747,461
6	Bengkulu	88,489
7	Kepulauan Bangka Belitung	2,289
8	Kepulauan Riau	226
9	Daerah Khusus Ibukota	0
10	Daerah Istimewa Yogyakarta	278,681
11	Banten	99,242
12	Bali	47,939
13	Nusa Tenggara Timur	665,009
14	Kalimantan Barat	179,712

15	Kalimantan Tengah	78,769
16	Kalimantan Selatan	305,685
17	Kalimantan Timur	70,752
18	Kalimantan Utara	3,181
19	Sulawesi Tengah	347,911
20	Sulawesi Tenggara	176,674
21	Sulawesi Barat	436,463
22	Maluku	24,895
23	Maluku Utara	124,445
24	Papua Barat	3,274
25	Papua	16,850

Sumber : Diolah dari data BPS (2017-2021)
Keterangan :
PJ (Produksi jagung)
PK (Pipilan Kering)

Meanwhile, there are eight provinces that are classified into the medium cluster as presented in Table 2. The range of dry shelled maize production in the medium cluster is between 1,008,794 to 2,914,504 tons of dry shelled maize.

Only one province is included in the high cluster, namely East Java, with an average dry shelled maize production of 5,427,877 per year. The average maize planting area during the 2017 – 2021 period in this region is the highest compared to other provinces, reaching 1,202,376 hectares.

Table 2. Medium cluster

No	Province	PJ(ton PK)
1	Sumatera Utara	1,442,790
2	Lampung	2,272,544
3	Jawa Barat	1,124,902
4	Jawa Tengah	2,914,504
5	Nusa Tenggara Barat	1,744,168
6	Sulawesi Utara	1,008,794
7	Sulawesi Selatan	1,862,630
8	Gorontalo	1,190,157

Sumber : diolah dari data BPS (2017-2021)

4. CONCLUSION

The K-Means Clustering algorithm with the Elbow method uses a dataset from 2017 to 2021 with annual data attributes of maize production (in dry shelled tons), maize harvested area (hectares), maize productivity (quintals per hectare), and maize planted area (hectares) produces three clusters of maize producing regions in Indonesia. There are 25 provinces in the low cluster, eight provinces in the medium cluster, and only one province in the high cluster, namely East Java.

REFERENCES

[1] Bastian, A., Sujadi, H. and Febrianto, G. (2018). Penerapan Algoritma K-Means Clustering Analysis pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka). J. Sist. Inf. (Journal Inf. Syst.,14(1), 26–32. https://doi.org/10.21609/jsi.v14i1.566.

[2] Fauzan, A., Baharudin, A.Y., and Wibowo, F. (2014). Sistem Klasterisasi Menggunakan Metode K-Means dalam Menentukan Posisi Access Point Berdasarkan Posisi Pengguna Hotspot di Universitas Muhammadiyah Purwokerto (Clustering System Using K-Means Method in Determining Access Point Position at Muhammadiyah University of Purwokerto). Jurnal Informatika (JUITA) 3(1), 25–29.

[3] Rahayu, S., Nugrahadi, D. T., and Indriani, F. (2014). Clustering Penentuan Potensi

ISSN: 2252-4274 (Print) 35
ISSN: 2252-5459 (Online)

Kejahatan Daerah Di Kota Banjarbaru Dengan Metode K-Means. Kumpulan Jurnal Ilmu Komputer (KLIK) 1(1), 33–45.

[4] Shirazy, A., Hezarkhani, A., Shirazi, A., Khakmardan, S., and Rooki, R. (2022). K-Means Clustering and General Regression Neural Network Methods for Copper Mineralization probability in Chahar-Farsakh, Iran. Geological Bulletin of Turkey 65(1), 79-92. https://doi.org/10.25288/tjb.1010636.

[5] Dedy, D., & Cherid, A. (2021). Data Mining Pengolahan Data Calon Pekerja Migran Indonesia (PMI) dengan Penerapan Metode Klustering K-Means dan Metode Klasifikasi K-Nearest Neighbor (KNN): Studi Kasus PT. SAM. *Format : Jurnal Ilmiah Teknik Informatika*, *9*(2), 166. https://doi.org/10.22441/format.2020.v9.i2.008

[6] G. Gustientiedina, M. H. Adiya, dan Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat- Obatan," Jurnal Nasional Teknologi dan Sistem Informasi, vol. 5, no. 1, hlm. 17–24, Apr 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.

[7] Tendean, T. and Purba, W. (2020). Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma K-Means. SAINTEK (Jurnal Sains dan Teknologi) 1(2), 5-11.

[8] Erlangga, N., Solikun, and Irawan. (2019). Penerapan Data Mining Dalam Mengelompokkan Produksi Jagung Menurut Provinsi Menggunakan Algoritma K-Means. KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer) 3(1), 702-709. DOI: 10.30865/komik.v3i1.1681

[9] Flomina G, K., Jihan Sy, Y., and Rahmawati, P. (2023). Klasterisasi Pembinaan Kelompok Tani Untuk Peningkatan Produksi Jagung Menggunakan Algoritma K-Means. Jurnal Teknika (Jurnal Fakultas Teknik Universitas Islam Lamongan) 15(1), 39-44. https://doi.org/10.30736/jt.v15i1.1012

[10] Hartati, Y., Defit, S., and Nurcahyo, G.W. (2021). Klasterisasi Bibit Terbaik Menggunakan Algoritma K-Means dalam Meningkatkan Penjualan. Jurnal Informatika Ekonomi Bisnis 3(1), 1-7. DOI: 10.37034/infeb.v3i1.56

[11] Silaban, E.F., Zunaidi, M. and Pane, D.H. (2020). Penerapan Data Mining Dalam Pengelompokan Bibit Jagung Unggul Menggunakan Algoritma K-Means. Jurnal CyberTech 3(2), 263-277

[12] Rosiani, U.D., Rahmad, C., Rahmawati, M.A., and Tupamahu, F. (2020). Segmentasi Berbasis K-Means Pada Deteksi Citra Penyakit Daun Tanaman Jagung. JIP (Jurnal Informatika Polinema) 6(3), 37-42.

.