

Division of Regions in Indonesia That Can Achieve Food Security for Corn

Romli Ardie^{1*}, Mukhtar^{2*}, Cecep Anwar Hadi Firdos Santosa³, Sholih⁴, Nana Hendracipta⁵

^{1,2,3,4,5}Universitas Sultan Ageng Tirtayasa

*Corresponding author. Email: ardietea@untirta.ac.id,
mukhtar@untirta.ac.id, cecepanwar@untirta.ac.id

Abstract - Clustering is a common technique for statistical data analysis, which is used in many fields, including machine learning, agriculture, and economics. Clustering method will be focused on clustering every province in Indonesia that is able to achieve food security of corn. In most developing countries, lack of resources, production and consumption are the main factors influencing food security in each province. Corn is an important agricultural commodity after rice because it can be used as a substitute for rice. Almost in some regions, corn still relies on staple food. This study aims to investigate the status of food security by classifying several provinces that can meet the needs of corn so that it will achieve food security. Data were obtained from 32 provinces, which included information on production and area of planting. From the results of clustering there are three categories of clustering. Silhouette Width is 0,73 that a good structure to the clusters.

Keywords: Clustering, hierarchical algorithm, food security, and R software

1. INTRODUCTION

Indonesia has a population of 225 million and is growing at a rate of 3 per year (BPS, 2017). Providing for food for citizen, is therefore, a real task and Indonesia must depend on imported food to guarantee food security. Agriculture is the main source of income and provides the core source of revenue for developing countries like Indonesia. Indonesia is an archipelagic country of about 17,000 islands. It is divided into 33 provinces, 501 districts, 6,694 sub-districts and 77,468 villages (BPS 2012). Its vast lands, 1.9 million km², are inhabited by about three hundred ethnicities. The country is well known for its distinctive culture and religious pluralism. The total population is

about 230 million, which ranks as the fourth largest in the world (BPS 2014).

Indonesia agriculture has been dominated by rice and corn. In the last 5 years, these crops contributed more than a one third in agriculture and occupied more than 30 percent of overall cultivated zone.

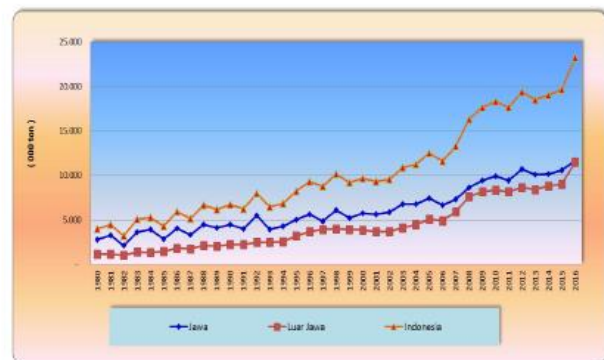


Figure 1.
Development of Corn Production in Indonesia, 1980-2016

Corn is one of the important food crop commodities after rice because in addition to being a source of calories or rice substitute food, corn is also used as animal feed. Corn needs will continue to increase from year to year in line with the improvement of the economic standard of living of the community and the progress of the animal feed industry so that efforts need to be made to increase production (Mandei, 2015).

The pattern of development of corn production in Java appears to be more fluctuating compared to the pattern of development of corn production in Outer Java. This happens because the competition of land use in Java, especially between rice and palawija can be the main reason for the

fluctuation of corn production level in Java, while corn production in Outer Java tends to increase slowly.

Corn production is obtained from the multiplication between harvested area and productivity. According to data from the Central Statistics Agency (2015c), the development of corn production in Indonesia in the period 1969 to 2014 tended to fluctuate but generally increased, similar to the pattern of harvested area, but the increase in production showed a more significant increase pattern (Ligawati, 2017).

Land area is an important component to increase corn production (Erviyana, 2014). This is due to the relationship between production and land area as production inputs (Badmus and Ariyo, 2011). Production is the relationship between production factors called inputs and production results called output (Sudarsono, 1984).

Cluster analysis is primarily focused on searching for similarities or differences among the examined objects. Cluster analysis provides one, empirically based, means for explicit classifying objects (Punjand Stewart, 1983). According to Everitt et al. (2011), cluster analysis techniques are concerned with exploring data sets to assess whether they can be summarized meaningfully in

terms of a relatively small number of groups or clusters of objects or individuals which resemble each other and which are different in some respects from individuals in other clusters (Majerova and Nevima, 2017).

In an effort to divide each province into manageable areas, each with a simple and representative form, including products and farming area of corn. The authors use cluster analysis to divide each province into several clusters.

The main objective of the presented paper is the analysis of current food security for corn in Indonesia at every province. This objective is accomplished by comparing province regarding analysis indicators between product and area of corn. In the second part are investigated province clustered into different groups according analysis indicators. Finally, created groups are compared and most different factors are identified.

2. METHOD AND MATERIALS

Dataset

The dataset used is the product and area farming of corn by province in Indonesia in 2015 which are available from BPS as in Table 1.

Table 1. Product and Area Corn

No	Province	Production	Harvest area
1	Aceh	205,125	47,967
2	North Sumatera	1,519,407	243,770
3	West Sumatera	602,549	87,825
4	Riau	30,870	12,425
5	Jambi	51,712	8,488
6	South Sumatera	289,007	46,298
7	Bengkulu	52,785	10,137
8	Lampung	1,502,800	293,521
18	East Nusa Tenggara	685,081	273,194
19	West Borneo	103,742	31,604
20	Central Borneo	8,189	2,506
21	South Borneo	128,505	21,926
22	East Borneo	8,379	2,307
23	North Borneo	1,032	474
24	North Sulawesi	300,490	80,872
25	Central Sulawesi	131,123	32,503

No	Province	Production	Harvest area
9	Bangka Belitung Islands	666	180
10	Riau Islands	473	203
11	West Java	959,933	126,828
12	Central Java	3,212,391	542,804
13	Yogyakarta	299,084	65,485
14	East Java	6,131,163	1,213,654
15	Banten	11,870	3,518
16	Bali	40,603	15,346
17	West Nusa Tenggara	959,973	143,117

No	Province	Production	Harvest area
26	South Sulawesi	1,528,414	295,115
27	North East Sulawesi	68,141	23,945
28	Gorontalo	643,512	129,131
29	West Sulawesi	100,811	20,752
30	Maluku	13,947	3,260
31	North Maluku	11,728	3,892
32	West Papua	2,264	1,273
33	Papua	6,666	2,495

From Table 1, we summarize statistics descriptively. Descriptive statistics are numbers that summarize the data with the

purpose of describing what occurred in the sample. Descriptive statistics are shown in Table 2.

Table 2. Descriptive statistics for production and land of corn in Indonesia.

	Min	Med	Mean	SD	Max
Product (Tons)	473	103,742	594,316.2	1,207,018	6,131,163
Land (Acres)	180	23,945	114,752	231,890.7	1,213,654

Descriptive statistics above the minimum value of production are found in the Riau is 473 ton and then the maximum is 6,131,163 in East Java. The smallest area in corn farming is in the province of Bangka Belitung Islands is 180 and then the maximum is 1,213,654 in West Sumatra.

It can be concluded that the most successful corn farming is in East Java and then the most extensive agricultural land is in the West Sumatra Province.

Analysis Cluster

Clustering is the process of grouping a set of objects in such a way that objects in the same group are more similar in some particular manner to each other than to those in other groups (Ghuman, 2016). Clustering plays a role of active research in several fields such as statistics, pattern recognition and

machine learning. Data mining adds complications to very large datasets with many attributes of different types to clustering. Clustering algorithms are used to organize and categorize data for data concretion and model construction, detection of deviation, etc. Common approach of clustering is to find centroid that will represent a certain cluster. Cluster centre will be represented with input vector which measures a similarity unit between input vector and all cluster centroid and determining which cluster is nearest or most similar one (Agrawal and Bojewar, 2014).

K-means

Given a dataset of n data points x_1, x_2, \dots, x_n such that each data point is in R^d , the problem of finding the minimum variance clustering of the dataset into k clusters is that

of finding k points $\{m_j\} (j = 1, 2, \dots, k)$ in R^d such that

$$\frac{1}{n} \sum_{i=1}^n [\min_j d^2(x_i, m_j)]$$

is minimized, where $d(x_i, m_j)$ denotes the Euclidean distance between

The K-means algorithm defined above aims at minimizing an **objective function**, which in this case is the **squared error** function. Suppose a data set, D , contains n objects in Euclidean space. Partitioning methods distribute the objects in D into K cluster, C_1, \dots, C_k , that is, $C_i \subset D$ and $C_i \cap C_j = \emptyset$ for $(1 \leq i, j \leq k)$.

To calculate the distance, we use Euclidean Distance as follows

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2}$$

Where

d_{ik} = distance of i -th data to center of cluster k

m = number of attribute

x_i = i -th data

c_k = centre of cluster k

In simpler terms, the objective function attempts to pick centroids that minimize the distance to all points belonging to its respective cluster so that the centroids are more symbolic of the surrounding cluster of data points (brilliant.org)

Silhouette

A silhouette coefficient value of an object, example A_j , is in the range between -1 to 1. The closer the object A_j to 1, the higher the degree of ownership of the object A_j in the cluster. While the value close to -1 (negative value) indicates that an object is likely to be on the wrong cluster, because the object similarity with other objects in the same cluster is low.

Silhouette coefficient test is a popular method that combines in a cluster cohesion

and separation between clusters. For each data point, the calculate the average involves three steps.

1. For each object i in the cluster, calculate the average distance from the object to the entire objects in its own cluster. Call the average distance between the object i to all members of the cluster as a_i .
2. For each object i and other cluster that is not cluster where i was in, calculate the average distance from object to the entire objects in other closest to cluster where i was (cluster neighbour). Call average distance to all objects in the neighboring cluster as b_i
3. For each object i , the silhouette coefficient obtained from:

$$\frac{(b_i - a_i)}{\max(a_i, b_i)}$$

The value of silhouette coefficient resides between -1 and 1. The negative value is not desirable because these values correlated with the case where the a_i , as the value of the average distance to other point in the cluster, larger than the value of b_i , the smallest average value to points on the other cluster. Silhouette coefficient value is expected to be positive ($a_i < b_i$) and a_i is expected to be as close as possible to 0, where the silhouette coefficient reaches its maximum value 1 when $a_i = 0$.

Table 3. SC Interpretation

SC Value	Interpretation
0.71 – 1.00	Strong Classification
0.51 – 0.70	Good Classification
0.26 – 0.50	Weak Classification
0 – 0.25	Bad Classification

The average value of the silhouette coefficient of the entire cluster will reflect the quality of the clusters. The average value of the Silhouette Coefficient (SC) is interpreted as a strong or weak classification as shown in table 3.

3. RESULTS AND DISCUSSION

Number of clusters

A fundamental question is how to determine the value of the parameter k (cluster). If we look the percentage of variance explained as a function of the number of clusters: one should choose a number of k (clusters) so that adding another cluster doesn't give much better modelling of the data. More precisely, if one plots the percentage of variance explained by the clusters against the number of clusters, the first clusters will add much information (explain a lot of variance), but some point the marginal gain will drop, giving an angle in the graph.

From clustering analysis, we get number of clusters $k = 3$. It means that $k = 3$ that the most optimal we choose three clusters and then proceed with cluster calculations using the k-means method. This calculation is based on between the variable corn production and the area of agricultural land. The sharp growth in distance when switching from a one to a tree-cluster solution occurs in almost all analyses and must not be viewed as a reliable indicator for the decision regarding the number of segments (Mooi and Sarstedt, 2011).

Further, based on the results of the calculation obtained among the first cluster with the central value (146,090.5; 35,692.54), the second cluster is at (1,613,819.7; 274,192.50) and the third is at (6,131,163; 1,213,654).

Every province is grouped among the first cluster, the second cluster, and the third cluster:

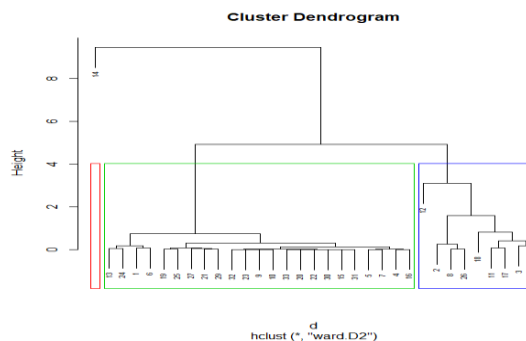


Figure 2. Cluster Dendrogram in every province of Indonesia

New method for analysis of clustering is based on the hierarchical clustering method. The dendrogram from R software is used to calculate and show the hierarchy of clusters. Dendrograms are tree diagrams that are a graphical representation of a hierarchical clustering of a data set.

The vertical axis of the dendrogram represents the distance or dissimilarity between clusters. The horizontal axis represents the clusters and objects between product and area of farming from corn.

Cluster I

From Figure 2, we have first cluster as following: Aceh, West Sumatera, Riau, Jambi, South Sumatera, Bengkulu, Bangka Belitung Island, Riau Island, Yogyakarta, Banten, Bali, East Nusa Tenggara, West Borneo, Central Borneo, South Borneo, East Borneo, North Borneo, North Sulawesi, Central Sulawesi, South East Sulawesi, West Sulawesi, Gorontalo, Maluku, North Maluku, West Papua, Papua.

Table 4. Descriptive statistics for first cluster

	Min	Med	Mean	SD	Max
Product (Tons)	473	52,248.5	146,090.5	206,395.9	685,081
Land (Acres)	180	13,885	535,692.5	58,424.5	273,194

Descriptive statistics for cluster I above the minimum value of production is found in the Riau is 473 and then the maximum is 685,081 in East Java. The smallest area in corn farming is in the

province of Bangka Belitung Islands is 180 and then the maximum is 273,194 in west Sumatra.

Cluster II

From Figure 2, we have second cluster as following: North Sumatera, Lampung, Jakarta, West Java, Central Java, West Nusa Tenggara. Descriptive statistics for cluster I (table 50 the minimum value of production is found in the Bengkulu is 52,785

tons island province and then the maximum is 3,212,391 in West Java. The smallest area in corn farming is in the province of Bengkulu is 10,137 and then the maximum is 542,804 in West Java.

Table 5. Descriptive statistics for second cluster

	Min	Med	Mean	SD	Max
Product (Tons)	52,785	1,239,690,0	1,372,150.5	1,050,059.7	3,212,391
Land (Acres)	10,137	193,443.5	226,961.8	183,713.6	542,804

Cluster III

In third cluster only the province of East Java between product is 6,131,163 tons and area are 1,213,654 acres. It is the biggest, among others.

Silhouette

Silhouette coefficient is a method of interpretation for the cluster validation on data objects. This technique provides a brief graphical representation of how well each object is located in its cluster.

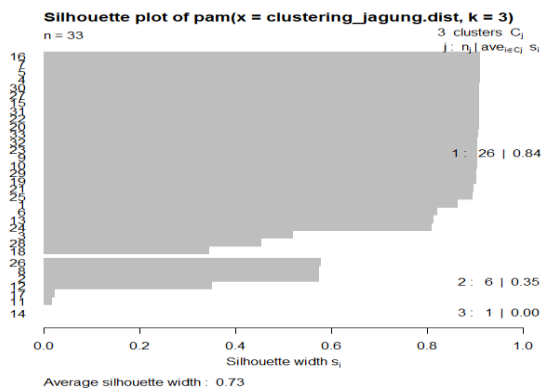


Figure 3 Silhouette

The plot indicates that there is a good structure to the clusters, with most observations seeming to belong to the cluster that they are in. There is a summary measure at the bottom of the plot label "Average Silhouette Width". This table shows how to use the value is 0,73.

4. CONCLUSION

This paper has showed how analysis of cluster analysis indicators between product

and area of corn to search for similarities across province and get better understanding about situation in every province in whole country. Cluster I has 26 Provinces (Aceh, West Sumatera, Riau, Jambi, South Sumatera, Bengkulu, Bangka Belitung Island, Riau Island, Yogyakarta, Banten, Bali, East Nusa Tenggara, West Borneo, Central Borneo, South Borneo, East Borneo, North Borneo, North Sulawesi, Central Sulawesi, South East Sulawesi, West Sulawesi, Gorontalo, Maluku, North Maluku, West Papua, Papua). Cluster II has 6 provinces (North Sumatera, Lampung, Jakarta, West Java, Central Java, West Nusa Tenggara), and cluster III has 1 province (East Java).

From clustering, we can see which region is potential to improve the corn production in Indonesia. Furthermore, the government must consider those clustering to make some policies regarding anticipate the product of corn in Indonesia for the future. East Java is the biggest corn production among others, while the population of east java is also the biggest in Indonesia with extremely limited land. On the other side, many provinces who have a bigger land and lower population have low corn production. By this condition, the government should have appropriate policies about production corn for the future.

REFERENCES

Ghuman, Sukhdev Singh. 2016. Clustering Techniques- A Review. International Journal of Computer Science and Mobile Computing: International

- Journal of Computer Science and Mobile Computing, Vol.5 Issue.5, May- 2016, pg. 524-530
- Agrawal, Akshay S and Bojewar, Sachin. 2014. Comparative Study of Various Clustering Techniques: , International Journal of Computer Science and Mobile Computing, Vol.3 Issue.10, October- 2014, pg. 497-504
- Lembang, Ferry Kondo; Yulita Lessil, Patresya; and Aulele, Salmon Notje. (2013). Pengelompokan Provinsi di Indonesia Berdasarkan PDRB atas Dasar Harga Konstan Tahun 2013: Jurnal Matematika Vol. 7, No. 2, Desember 2017, pp. 76-84
- Majerova, Ingrid and Nevima, Jan. (2017). The measurement of human development using the Ward method of cluster analysis: Journal of International Studies, 10(2), 239-257. doi:10.14254/2071-8330.2017/10-2/1
- Badmus., and O.S. Ariyo. (2011). Forecasting Cultivated Areas and Production of Maize in Nigerian using ARIMA Model M.A. Asian Journal of Agricultural Sciences 3(3): 171-176, 2011
- Erviyana, Poppy. (2014). Faktor- Faktor yang Mempengaruhi Produksi Tanaman Pangan Jagung di Indonesia. Journal of Economics and Policy: Jejak 7 (2) (2014): 100-202. DOI: 10.15294/jejak.v7i1.3596
- Punj, G., & Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. Journal of marketing research, 134-148.
- Ligawati, Linda. (2017). Analisis Produksi dan Konsumsi Jagung Domestik dalam Rangka Pencapaian Swasembada Jagung Nasional tahun 2017. /repository.ipb.ac.id
- Sudarsono. (1984). dalam Joko Triyanto. (2006). Analisis Produksi Padi Di Jawa Tengah. Semarang: Universitas Diponegoro
- Everitt, B.S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster analysis. Chichester: John Wiley & Sons
- Mandei, Juliana Ruth.(2015). ASE – Volume 11 Nomor 1, Januari 2015: 28 - 37
- Oyelade, O. J et all. 2010. Application of k-Means Clustering algorithm for prediction of Students' Academic Performance: (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, o. 1, 2010
- <https://brilliant.org/wiki/k-means-clustering/>