International Journal of Information System & Technology
Akreditasi No. 36/E/KPT/2019 | Vol. 3, No. 2, (2020), pp 276-283

GRDP Growth Rate Clustering in Surabaya City uses the K-Means Algorithm

Nur Ahlina Febriyati [1], *Achmad Daengs GS* [2], *Anjar Wanto* [3]
[1,2] *Universitas 45 Surabaya, Surabaya, Indonesia*
[3] *STIKOM Tunas Bangsa, Pematangsiantar, Indonesia*

[*] *anjarwanto@amiktunasbangsa.ac.id*

Abstract

Gross Regional Domestic Product (GRDP) is an indicator used to measure the economic performance of a region in a period. GRDP is the amount of added value generated by all business units in a particular area and can also be said to be the sum of the value of the final goods and services produced by all economic units. Therefore, this study aims to cluster the GRDP Growth Rate according to business fields in the city of Surabaya, so that it is known which sectors have high or low growth. The clustering algorithm used is K-Means. By using this method the data will be grouped into several clusters, where the implementation of the K-Means Clustering process uses the Rapid Miner tools. The data used is the GRDP Growth Rate in Surabaya City by Business Field, 2010-2019 (Percent). The data is divided into 3 clusters: high, medium and low. The results obtained are that there are 9 categories / sectors with high clusters, 5 categories / sectors with medium clusters and 3 categories / sectors with low clusters. This can be input and information for the Surabaya City government to further maximize efforts to increase the GRDP Growth Rate in the area.

Keywords: *Clustering, Growth Rate, GRDP, Surabaya City, K-Means*

1. Introduction

GRDP is the amount of added value generated by all business units in a certain area. It can also be said to be the sum of the value of the final goods and services produced by all economic units. [1] [2]. Gross Regional Domestic Product (GRDP) is an indicator used to measure the economic performance of a region in a period [3][4]. The benefits of calculating the GRDP value are: (1) knowing and analyzing the structure or composition of the economy. From the GRDP calculation, it can be seen whether an area is an industrial, agricultural or service area and how much is the contribution of each sector; and (2) comparing the economy over time. Because the GRDP value is recorded every year, it will get a record number from year to year. Thus, it is hoped that information on the increase or decrease in whether there is a change or reduction in material prosperity or not [5]. For Indonesia, which is a developing country, economic growth is one of the goals that must be achieved in the implementation of development. The achievements of the Indonesian economy are strongly influenced by the existence of supporting sectors as part of economic improvement.

The specific purpose of this research is to provide input and information for the Surabaya City government to further maximize efforts to increase the GRDP Growth Rate in the area. Research data in the form of GRDP Growth Rate by Business Field in the City of Surabaya were obtained from the Central Statistics Agency of Surabaya City. The grouping method used in this study is the K-Means Clustering data mining algorithm. Apart from grouping, data mining is also often used for data classification problems [6]–[10].

Table 1. GRDP Growth Rate by Business Field, 2017-2019 (Percent)

No	Category / Sector	2017	2018	2019
1	Agriculture, Forestry and Fisheries	3,35	-1,44	-0,90

ISSN : 2580-7250

No	Category / Sector	2017	2018	2019
2	Mining and excavation	2,58	1,06	0,09
3	Processing industry	4,95	4,93	5,43
4	Procurement of Electricity and Gas	1,75	-0,07	0,41
5	Water Supply, Waste Management, Waste and Recycling	6,83	3,32	3,92
6	Construction	6,92	6,28	5,45
7	Wholesale and Retail Trade; Car and Motorcycle Repair	6,01	6,36	5,89
8	Transportation and Warehousing	6,87	7,56	7,62
9	Provision of Accommodation and Food and Drink	8,64	7,80	7,67
10	Information and Communication	6,93	6,83	7,49
11	Financial Services and Insurance	2,92	4,84	3,49
12	Real Estate	5,08	5,78	6,09
13	Company Services	6,66	7,12	6,84
14	Mandatory Government Administration, Defense and Social Security	4,85	2,17	3,64
15	Education Services	5,95	6,24	6,66
16	Health Services and Social Activities	5,87	7,81	7,64
17	Other services	2,89	5,18	5,94

Source: Central Bureau of Statistics for the City of Surabaya [11]

Many previous studies are related to this research, which discusses grouping using the K-Means algorithm, including: Research on the grouping of rice plants in Indonesia based on 34 provinces. The results of this study were in the form of rice plant data grouping which were divided into 3 clusters, namely the high cluster consisting of 3 provinces, the normal cluster consisting of 23 provinces and the low cluster consisting of 8 provinces [12]. Subsequent research was carried out for grouping disaster-prone areas based on provinces in Indonesia. The results of this study are data grouping of disaster-prone areas divided into 3 clusters, namely the high cluster consisting of 4 provinces, the normal cluster consisting of 14 provinces and the low cluster consisting of 16 provinces [13]. Next, the research was conducted to classify population density, human development index, open unemployment rate and school enrollment rates by province in Indonesia. The results of this research are cluster 1 consisting of 12 provinces, cluster 2 consisting of 6 provinces and cluster 3 consisting of 1 province, cluster 4 consisting of 6 provinces and cluster 5 consisting of 9 provinces [14]. These related studies are the background for conducting research to cluster the GRDP Growth Rate by Business Field in the city of Surabaya. The results of this study are in the form of clustering information on GRDP Growth Rate according to the Business Field in the city of Surabaya which is expected to provide input and information for the Surabaya City government to further maximize efforts to increase the GRDP Growth Rate in the area.

2. Research Methodology

2.1. Method of collecting data

Data collection uses quantitative methods, namely GRDP Growth Rate data according to business fields in the city of Surabaya in 2017-2019 (percent) taken from the Central Statistics Agency of Surabaya city.

2.2. Research Flowchart

The research flowchart of the K-Means clustering algorithm is presented in Figure 1 [15][16].

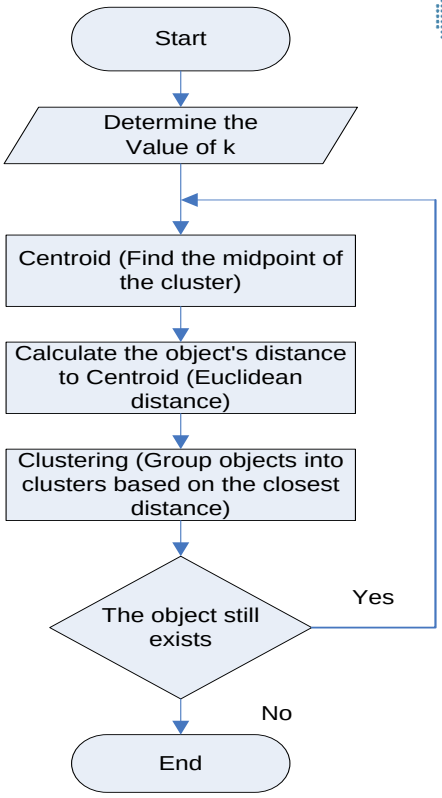

Figure 1. Research Flowchart

The steps of the K-Means algorithm can be explained as follows [17]–[21]:

1. Determine the number of clusters (k) in the data set.
2. Determine the center value (Centroid)
   Determination of the centroid value at the initial stage is carried out randomly, while in the iteration stage the formula is used as in equation (1) below.

$$V_{ij} = \frac{1}{Ni} \quad (1)$$

   Explanation :
   V_{ij} = Centroid cluster average to-I for variable to-j
   N_i = Number of cluster members to -i
   i, *k* = Index of cluster
   j = Index of variable
   Xkj = data vakue to-k variable to-j for that cluster
3. On each record, calculate the closest distance to Centroid.
   There are several ways that can be used to measure the distance of data to the center of the group, including Euclidean, Manhattan / City Block, and Minkowsky. Each method has advantages and disadvantages of each. For writing in this chapter, the Centroid distance used is Euclidean Distance, with the following formula.

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2} \quad (2)$$

   Explanation:
   De = Euclidean Distance
   i = The number of objects [2]
   (x, y) = Object coordinates
   (s, t) = Centroid coordinates
4. Group objects by distance to the nearest Centroid
5. Repeat step 3 to step 4, iterating until Centroid is optimal.

3. Results and Discussion

3.1. Data Normalization

Before the data is processed using the K-Means algorithm, the data must first be normalized by taking the average GRDP Growth Rate according to business fields in the city of Surabaya (based on table 1). So the normalization data is obtained as in table 2.

Table 2. Data Normalization

No	Category / Sector	Average
1	Agriculture, Forestry and Fisheries	0,34
2	Mining and excavation	0,70
3	Processing industry	1,24
4	Procurement of Electricity and Gas	3,55
5	Water Supply, Waste Management, Waste and Recycling	3,75
6	Construction	4,67
7	Wholesale and Retail Trade; Car and Motorcycle Repair	4,69
8	Transportation and Warehousing	5,10
9	Provision of Accommodation and Food and Drink	5,65
10	Information and Communication	6,09
11	Financial Services and Insurance	6,22
12	Real Estate	6,28
13	Company Services	6,87
14	Mandatory Government Administration, Defense and Social Security	7,08
15	Education Services	7,11
16	Health Services and Social Activities	7,35
17	Other services	8,04

3.2. Centroid Data

To get the midpoint value on Centroid data, it is necessary to make a provision that the desired clusterization is 3. Determination of the cluster is divided into 3 parts, namely the high cluster (C1), medium cluster (C2) and low cluster (C3). For the cluster point value is determined by taking the maximum value for the high cluster (C1), the average value for the medium cluster (C2) and the minimum value for the low-level cluster (C2). The cluster point value can be seen in table 3 (Based on table 1).

Table 3. Initial Centroid Value

Centroid	C1 (Max)	8,64	7,81	7,67
	C2 (Ave)	5,24	4,81	4,90
	C3 (Min)	1,75	-1,44	-0,90

3.3. K-Means clustering with Rapidminer

The use of the K-Means algorithm for clustering the GRDP Growth Rate by Business Field in the City of Surabaya with the help of the RapidMiner 5.3 software is divided into 3 clusters, namely: high, medium and low. Following are the results of clustering and using the K-Meanss model with Rapid Miner 5.3.

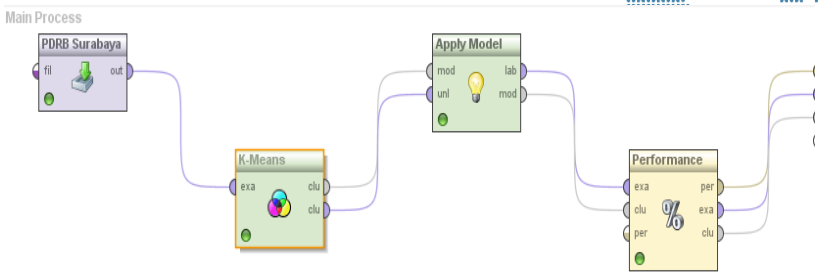

Figure 2. Process K-Means with Rapidminer (K Value = 5)

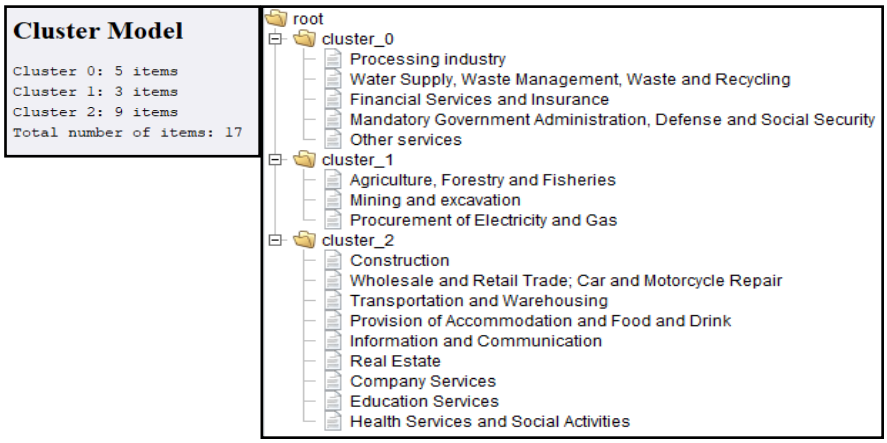

Figure 3. Model Cluster and Results Clustering using K-Means

Based on Figure 3 it can be explained that of the three resulting clusters there are 5 items for Cluster_0, 3 items for Cluster_1 and 9 items for Cluster_2. For the final result of the Centroid table can be seen in Figure 4. While the results of the plot view cluster GRDP Growth Rate by Business Field in the City of Surabaya is presented in Figure 5.

Attribute	cluster_0	cluster_1	cluster_2
2017.0	4.488	2.560	6.548
2018.0	4.088	-0.150	6.864
2019.0	4.484	-0.133	6.817

Figure 4. Centroid table

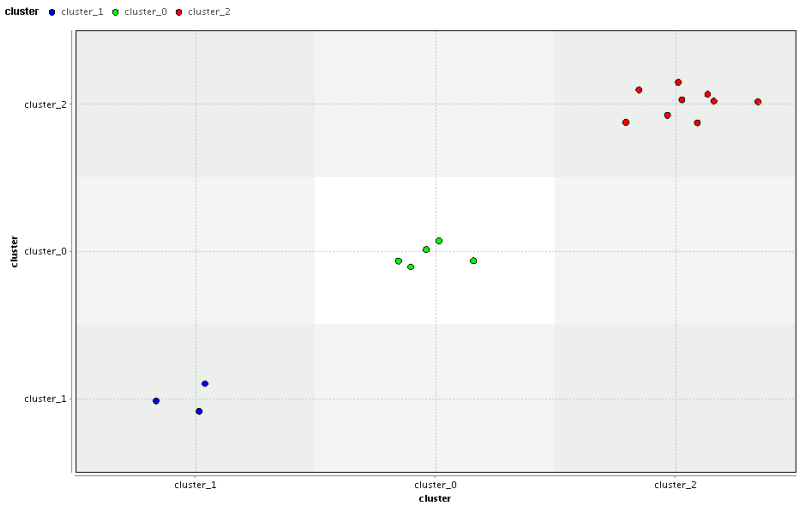

Figure 5. Plot View of Surabaya City GRDP Growth Rate

Based on Figure 3 and Figure 4 and Figure 5 it can be explained that the low cluster is in Cluster_1 which consists of 3 items whose description can be seen in Figure 3. The cluster is currently at Cluster_0 which consists of 5 items whose description can also be seen in Figure 3 and the cluster height is at Cluster_2 which consists of 9 items (the description can be seen in figure 3).

3.4. Performance Vector K-Means

Seeing the performance of K-Means using the Rapid Miner tools is to add the% Performance operator which can evaluate the performance of the centroid-based clustering algorithm. This operator provides a list of performance criteria values based on cluster centroid. % Performance measurement parameters are Avg._within_centroid_distance and Davies Bouldin. In this study the results of Avg._within_centroid_distance = -2721 and Davies_Bouldin = -0,712. The results of K-Means performance using the Rapid Miner tools can be seen in Figure 6.

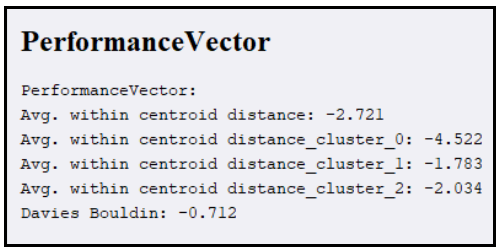

Figure 6. Performance Vector K-Means dengan Rapidminer

4. Conclusion

The use of the K-means algorithm can be used for clustering the GRDP Growth Rate by Business Field in Surabaya City according to business fields. The results of the research are in the form of clusters consisting of 3 clusters, including: Cluster_1 (1. Agriculture, Forestry and Fisheries, 2. Procurement of Electricity and Gas, 3. Mining and excavation). Cluster_0 (1. Mandatory Government Administration, Defense and Social Security, 2. Financial Services and Insurance, 3. Other services, 4. Water Supply, Waste Management, Waste and Recycling, 5 Processing industry). Cluster_2 (1. Real Estate, 2. Wholesale and Retail Trade; Car and Motorcycle Repair, 3. Construction, 4. Education Services, 5. Company Services, 6. Information and Communication, 7. Health Services and Social Activities, 8. Transportation and Warehousing, 9. Provision of Accommodation and Food and Drink).

References

[1] M. Muliza, "Analisis Pengaruh Jumlah Penduduk Dan Produk Domestik Regional Bruto (GRDP) Terhadap Kemiskinan Di Kabupaten Nagan Raya," *At-Tasyri': Jurnal Ilmiah Prodi Muamalah*, vol. 12, no. 1, p. 63, 2020.

[2] S. M. AZ and S. Pebrina, "Analisis Pengaruh Tingkat Investasi, Pendapatan Asli Daerah dan Tenaga Kerja Terhadap GRDP Provinsi Jambi," *Ekonomis: Journal of Economics and Business*, vol. 3, no. 1, pp. 36–43, 2019.

[3] A. B. M. Bintang and N. Woyanti, "Pengaruh GRDP, Pendidikan, Kesehatan, Dan Pengangguran Terhadap Tingkat Kemiskinan di Jawa Tengah (2011-2015)," *Media Ekonomi dan Manajemen*, vol. 33, no. 1, pp. 20–28, 2018.

[4] R. Giovanni, "Analisis Pengaruh GRDP, Pengangguran dan Pendidikan Terhadap Tingkat Kemiskinan di Pulau Jawa Tahun 2009-2016," *Economics Development Analysis Journal*, vol. 7, no. 1, pp. 23–31, 2018.

[5] I. Ummi, Rusdarti, and H. Yanto, "Relationship of the GRDP Sectors with Environmental Quality Index in Indonesia 2012-2017," *Journal of Economic Education*, vol. 8, no. 2, pp. 152–158, 2019.

[6] I. S. Damanik, A. P. Windarto, A. Wanto, Poningsih, S. R. Andani, and W. Saputra, "Decision Tree Optimization in C4.5 Algorithm Using Genetic Algorithm," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, Aug. 2019.

[7] W. Katrina, H. J. Damanik, F. Parhusip, D. Hartama, A. P. Windarto, and A. Wanto, "C.45 Classification Rules Model for Determining Students Level of Understanding of the Subject," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.

[8] H. Siahaan, H. Mawengkang, S. Efendi, A. Wanto, and A. Perdana Windarto, "Application of Classification Method C4.5 on Selection of Exemplary Teachers," *Journal of Physics: Conference Series*, vol. 1235, no. 1, pp. 1–7, Jun. 2019.

[9] I. Parlina *et al.*, "Naive Bayes Algorithm Analysis to Determine the Percentage Level of visitors the Most Dominant Zoo Visit by Age Category," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–5, 2019.

[10] D. Hartama, A. Perdana Windarto, and A. Wanto, "The Application of Data Mining in Determining Patterns of Interest of High School Graduates," *Journal of Physics: Conference Series*, vol. 1339, no. 1, pp. 1–6, 2019.

[11] B. P. Statistik, "Laju Pertumbuhan GRDP Menurut Lapangan Usaha 2010-2019 (Persen) 2017-2019," *Badan Pusat Statistik Kota Surabaya*, 2020. [Online]. Available: https://surabayakota.bps.go.id/indicator/156/204/1/laju-pertumbuhan-GRDP-menurut-lapangan-usaha-2010-2019.html. [Accessed: 10-Apr-2020].

[12] Sudirman, A. P. Windarto, and A. Wanto, "Data mining tools | rapidminer: K-means method on clustering of rice crops by province as efforts to stabilize food crops in Indonesia," *IOP Conference Series: Materials Science and Engineering*, vol. 420, no. 1, pp. 1–8, 2018.

[13] B. Supriyadi, A. P. Windarto, T. Soemartono, and Mungad, "Classification of Natural Disaster Prone Areas in Indonesia using K-Means," *International Journal of Grid and Distributed Computing*, vol. 11, no. 8, pp. 87–98, 2018.

[14] A. S. Ahmar, D. Napitupulu, R. Rahim, R. Hidayat, Y. Sonatha, and M. Azmi, "Using K-Means Clustering to Cluster Provinces in Indonesia," *Journal of Physics: Conference Series*, vol. 1028, no. 1, pp. 1–6, 2018.

[15] Z. S. Younus *et al.*, "Content-based image retrieval using PSO and k-means clustering algorithm," *Arabian Journal of Geosciences*, vol. 8, no. 8, pp. 6211–6224, 2015.

[16] A. Wanto *et al.*, *Data Mining : Algoritma dan Implementasi*. Yayasan Kita Menulis, 2020.

[17] E. Prasetyo, *Data Mining: Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta: Andi Offset, 2012.

[18] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*. New Jersey: John Wiley & Sons, 2005.

[19] R. Primartha, *Belajar Machine Learning Teori dan Praktik*. Bandung: Informatika Bandung, 2018.

[20] T. Khotimah, "Pengelompokan Surat dalam Al Qur'an menggunakan Algortima K-Means," *Jurnal Simetris*, vol. 5, no. 1, pp. 83–88, 2014.

[21] I. Parlina, A. P. Windarto, A. Wanto, and M. R. Lubis, "Memanfaatkan Algoritma K-Means dalam Menentukan Pegawai yang Layak Mengikuti Asessment Center untuk Clustering Program SDP," *CESS (Journal of Computer Engineering System and Science)*, vol. 3, no. 1, pp. 87–93, 2018.

Authors

1st Author
Nur Ahlina Febriyati
Lecturer of Universitas 45 Surabaya, Surabaya, Indonesia
nurahlinaf@yahoo.co.id

2nd Author
Achmad Daengs GS
Lecturer of Universitas 45 Surabaya, Surabaya, Indonesia
bumigora80@gmail.com

3rd Author
Anjar Wanto
Lecturer of STIKOM Tunas Bangsa, Pematangsiantar, Indonesia.
anjarwanto@amiktunasbangsa.ac.id