Analysis of District Cluster Based on the Indicator of Gross Regional Domestic Product (GRDP) Using Unsupervised Learning

Agung Jaelanie
Dept of Electrical Engineering
Faculty of Industrial Technology
Institute Technology SepuluhNopember
agung14@mhs.ee.its.ac.id

Supeno Mardi Susiki Nugroho
Department of Electrical Engineering
Faculty of Industrial Technology
Institute Technology SepuluhNopember
mardi@its.ac.id

Mochamad Hariadi
Department of Electrical Engineering
Faculty of Industrial Technology
Institute Technology SepuluhNopember
mochar@ee.its.ac.id

Fachrul Kurniawan
Department of Electrical Engineering
Faculty of Industrial Technology
Institute Technology SepuluhNopember
fachrulkurniawan873@gmail.com

Abstract- The economic lameness among regions is an unsolved problem all this time. To determine policies, to plan economic development right on target, and to solve the economic lameness, we need sufficient and accurate data and information. In this research, there's data of Gross Regional Domestic Product (GRDP) would be analyzed here to get the economic pattern of District cluster in Indonesia. The tool for this analysis in this research is clustering technic in which it has a method called K-Means. K-means groups the obtained data based on its features for each. This research results in cluster 5 as the optimal ones. With the value of DBI 0.9640.

Keywords: Clustering, K-Means Algorithm, Davies-Bouldin Index (DBI), Gross Regional Domestic Product (GRDP).

INTRODUCTION

Economic development is a series of efforts aimed to improve public welfare, to minimize the imbalance of revenue in society. Economic development will run effectively and efficiently when it rests on the plan of development right on target. Besides, the evaluation of policy is required to fix the past planning of development and futures. Gross Regional Domestic Product (GRDP) is one of the indicators that used to measure the economic activities in certain region. Indonesia is archipelago state that every region in Indonesia has its economic characteristic.

The purpose of this research is to know every pattern of the economy and dominant sector in a certain region based on GRDP. The search of a pattern is implemented by clustering approach. The method of this research is K-Means. K-Means Algorithm is a method of grouping with unsupervised learning[4]. Cluster evaluation is finished by validating Davies-Bouldin Index (DBI) to determine the numbers of optimal cluster.

This research takes cluster 5 as an optimal cluster, in which cluster 1 consists of 3 districts, cluster 2 consists of 15 districts, cluster 3 consists of 36 districts, cluster 4 consists of 6 districts, and cluster 5 consists of 388 districts. Cluster 5 is a district that has lower economic level than other clusters.

This information of economic pattern expected to be a subject of related agency in time of determining policy, planning activities, making a program of welfare improvement and minimizing the imbalance of revenue among regions proportionally, accurately and repeatedly until the public welfare is well achieved.

METHODOLOGY

This research has 2 targets. The first target is to decide the numbers of optimal-shaped cluster. The second target is to do analysis of economic pattern characteristic in every cluster. To acquire more information, several techniques are used as follows:

- Research test is applied by giving the numbers of the cluster from 3 to 7 clusters with K-Means method.
- Decide the numbers of optimal clusters by using validated Davies-Bouldin Index (DBI).
- The analysis in every cluster should be done with visualization of the parallel coordinate. It is aimed to know the characteristic of economic pattern in every cluster.

RESULTS AND DISCUSSION

TABLE 1. VALUE OF DBI

Number of Cluster	3	4	5	6	7
DBI	1.1015	1.0595	0.9640	0.9683	1.0603

This research results in cluster 5 as the optimal ones. With the value of DBI 0.9640. Cluster 1 with 3 districts, cluster 2 with 15 districts, cluster 3 with 36 districts, cluster 4 with 6 districts, and cluster 5 with 388 districts. Results of clusters centroid show in Table 2.

TABLE 2. CENTROID OF CLUSTER

Attribute	C1	C2	C3	C4	C5
PRT	0.4058	0.0935	0.4724	0.0131	0.0788
PTM	0.8255	0.0068	0.0474	0.0000	0.0075
IP	0.0359	0.2921	0.0637	0.2906	0.0104
PLG	0.0029	0.2845	0.0119	0.2365	0.0025
PAR	0.0296	0.2967	0.0317	0.3530	0.0166
K	0.0837	0.2550	0.0625	0.8478	0.0180
PER	0.0372	0.1639	0.0473	0.6740	0.0131
TRA	0.0410	0.2920	0.0452	0.6473	0.0251
PAM	0.0046	0.0500	0.0116	0.4501	0.0054
IK	0.0097	0.1060	0.0228	0.5900	0.0078
JK	0.0027	0.0398	0.0058	0.3411	0.0029
RE	0.0119	0.0902	0.0173	0.6333	0.0084
JP	0.0010	0.0162	0.0021	0.5656	0.0012
AP	0.0295	0.0508	0.0235	0.4063	0.0123
JPEN	0.0296	0.1066	0.0323	0.6976	0.0159
JKES	0.0336	0.1113	0.0247	0.6963	0.0174
JL	0.0083	0.0709	0.0177	0.5314	0.0067

Cluster one is leading more in agriculture and mining sector with score 0.4058 and 0.8255 than other sectors with low scores. Cluster 2 is leading in the processing industry, electricity and gas supply, water supply, construction and transportation sector with score 0.2921, 0.2845, 0.2967, 0.255 and 0.2920 compared to other sectors with low scores. Cluster 4 is leading in construction, business, transportation, real estate, company services, office administration, education services, health services, and other services with score 0.8478, 0.6740, 0.6473, 0.6333, 0.5656, 0.4063, 0.6976, 0.6963 and 0.5314, but cluster's agriculture and mining sector is very low with score 0.0131 and 0.0000. Cluster 5 prevails every sector with score under 0.0788. There is no dominant sector in cluster 5.

CONCLUSION

From the result of cluster district based on GRDP, it can be concluded that K-Means method can be used to undertake to cluster and to know the pattern of district economy. Cluster 1 dominated by a district/city in East Kalimantan and in Riau. Both islands are the largest mining producers. Cluster 2 dominated by a district in West Java, East Java, North Sumatera, and South Sumatera. This region consists of many industrial areas. Cluster 3 dominated by the district in West Java, East Java, Riau and Lampung. These areas are the largest agriculture producers. Cluster 4 dominated by Jakarta and Surabaya city because this area is the center of economic movement in Indonesia. Cluster 5 dominated by some districts in east of Indonesia like Papua, NTT, NTB, Maluku, a half of Sulawesi and the other half in middle of Kalimantan. There is no sector as dominant as in these areas. Cluster 5 needs special concern from the government since their economy is fairly low compared to other areas in Indonesia.

REFERENCES

[1] BPS., Produk Domestik Regional Bruto Kabupaten/Kota Menurut Lapangan Usaha, Publikasi, 2015.

(BPS., The domestic product of Gross regional district/city viewed in the business field, Publication 2015)

[2] Sri, M., Edi, W., "*Teknik Visualisasi Dalam Data Mining*", Seminar Nasional Informatika (semnasIF), UPN Veteran, Yogyakarta, 2009. ("The technique of Visualization in data of Mining," National Informatics Seminar (semnasIF), UPN Veteran, Jogjakarta)

[3] Kiem A. Daniel, "*Information Visualization and Visual Data Mining*", IEEE Transactions Visualizations and Computer Graphics, Vol.7, No.1, January-March, 2002.

[4] JinHuaXu, HongLiu, "*Web User Clustering Analysis on KMeans Algorithm*", International Conference on Information, Networking and Automation (ICINA), 2010.

[5] N. Hadi., A. Marzieh., K. Manijeh., F. S. Vahid., "*The Impact of Distance Metrics on K-means Clustering Algorithm Using in Network Intrusion Detection Data*", International Journal of Computer Networks and Communications Security, Vol.3, No.5, 225–228, May, 2015.

[6] B. J. Dibya., Dr. G. K. Anil., "*Effect of Different Distance Measures on the Performance of K-Means Algorithm: An Experimental Study in Matlab*", International Journal of Computer Science and Information Technologies (IJCSIT), Vol.5 (2), 2501-2506, 2014.

[7] Davies, D. L., Bouldin, D. W., "*A Cluster Separation Measure*", IEEE Transactions on Pattern Analysis and Machine Intelligence (2): 224, 1979.