Jurnal Varian
Vol. 6, No. 1, Oktober 2022, Hal. 25–34
e-ISSN: 2581-2017

Clustrering of BPJS National Health Insurance Participant Using DBSCAN Algorithm

Wiwit Pura Nurmayanti[1], Dewi Juliah Ratnaningsih[2], Sausan Nisrina[3], Abdul Rahim[4], Muhammad Malthuf[5], Wirajaya Kusuma[6]

[1,3]Departement of Statistics, Faculty of Mathematics and Natural Science, Universitas Hamzanwadi, Indonesia
[2]Departement of Statistics, Faculty of Science and Technology, Universitas Terbuka, Indonesia
[4]Departement of Pharmacy, Faculty of Pharmacy, Universitas Mulawarman, Indonesia
[5]5Departemen of Social Education, Faculty of Teacher Training and Education, Mataram State Islamic University, Indonesia
[6]Departement of Computer Science, Faculty of Engineering, Universitas Bumigora, Indonesia

Article Info

Article history:

Received : 04-02-2022
Revised : 09-08-2022
Accepted : 11-13-2022

Keywords:

Clustering;
DBSCAN;
JKN BPJS Health;
Noise;
Outlier;
Spatial.

ABSTRACT

In the current era of Big Data, getting data is no longer a difficult thing because they can access easily it via the internet, which is open access. A large amount of data can cause many problems in the data, such as data that deviates too far from the average (outliers). The method used to handle outlier data is DBSCAN which is density based clustering. The DBSCAN can be applied in various fields, one of which is the social sector, namely the participation of the JKN BPJS Health in West Nusa Tenggara. This study sees the distribution of BPJS Health participation groups, and to detect outliers so that objects with noise are not included in the cluster. The results of the study using the DBSCAN algorithm show that the optimal epsilon value is between 0.37 points by observing the knee of a curve. and MinPts 3, with the highest silhouette value of 0.2763. The highest JKN BPJS participants are in cluster 1 with 5 sub-districts, the second highest cluster is cluster 3 with 5 sub-districts, while the lowest cluster is cluster 2 with 93 sub-districts. The 13 sub-districts are not included in any group because they are noise data.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v6i1.1886

Corresponding Author:

Wiwit Pura Nurmayanti
Department of statistics, Universitas Hamzanwadi
Email: wiwit.adiwinata3@gmail.com

A. INTRODUCTION

Currently, data plays a very important role because data contains various kinds of information that can be used for various purposes, such as the basis for deciding or policies, predicting or forecasting the future, customer behavior and many others. In the current era of Big Data, getting data is no longer a difficult thing because they can access easily it via the internet such as the website of a government agency or industry because of its open access nature. However, with conditions like this, of course, there are not a few demands to find useful information that is drowned in piles of data from various sources. The large amount of data will be very difficult if you want to analyze whether there is an error in the data (Fisher et al., 2012) (Matuschek et al., 2017). Data that deviates too far from other data in a data set is called outlier (Aggarwal, 2017) (Zimek and Filzmoser, 2018). If there are data outliers, the results of data analysis will be biased, or do not reflect the real problem. However, outlier data sometimes have very useful information, which is defined as abnormal system data (Fisher et al., 2012) (Gebremeskel et al., 2016) (Zimek and Filzmoser, 2018).

The data mining methods in dealing with outlier data is clustering, where clustering defines an outlier that is not in the cluster. Clustering implicitly defines outliers as noise from a particular cluster. In clustering, there is a method that can detect outliers, namely Density Based Spatial Clustering Algorithm with Noise (DBSCAN) by performing several additional computations (Campello et al., 2015) (Kha and Anh, 2015) (Malkomes et al., 2015). DBSCAN is a grouping algorithm based on areas that have a higher data density than the surrounding area. The DBSCAN method was developed based on the density algorithm so that it grows high enough areas into clusters and finds these clusters in an arbitrary form in a spatial database containing noise (Galán, 2019) (Zhao et al., 2015). The DBSCAN method is used because it can overcome the weakness of data that has outliers (Pourbahrami et al., 2020).

Clustering can occur by grouping various data (Hendayanti et al., 2018) (Saky et al., 2020), one of which can group sub-districts based on the number of health insurance ownership in NTB Province. The National Health Insurance (JKN) grouping aims to identify JKN BPJS participants, as well as detect JKN participants who are still low or far from the average, especially BPJS PBI participants who are poor people. According to (BPS NTB, 2021) health is an outcome of development, because progress in development allows the availability of proper health facilities for the community, in the economic field increasing peoples income so that they can access higher quality health services. Access and health services have become a common problem, so the government seeks to ensure easy access and health services for all people through JKN. On this basis, the authors tested the performance of the DBSCAN method based on data that has certain characteristics, so that it will get an effective and optimal data grouping.

Previous research related to JKN BPJS clustering has been carried out by (Sadewo et al., 2021) to cluster BPJS health sentiment data using the AHC Average Linkage algorithm and the results got are negative data giving optimal results on the number of clusters 8 and intersection points 683, and the level of accuracy of the AHC Average Linkage method tested with the silhouette coefficient on negative data is the highest average value of 0.9953. Another study was conducted by (Herlinda et al., 2021), which applied the Fuzzy C-Means (FCM) method to classify the profile mapping data of health care providers. The result of their research is that the application of FCM resulted in 2 clusters, in the first cluster 479 and in the second cluster 580. A similar study was also conducted by (Ali and Masyfufah, 2021), they clustered BPJS patients using the K-Means Clustering method and the results got were 3 clusters, the first cluster comprised 91 female patients (51%), in the second cluster was 26 female patients (14%), and the third cluster comprised 63 male patients (35%). Patients who suffer the most pain from Krian and Balonbendo sub-districts, for patients with the female gender, the disease is more dominant than the male gender. This study complements previous studies. In previous studies related to BPJS clustering, many used familiar clustering and could not handle data that had deviations, such as outlier data. In contrast to this study, in performing clustering using DBSCAN this method can detect outliers by performing several additional computations, while the AHC Average Linkage, Fuzzy C-Means and K-Means Clustering methods could not perform clustering on data that has outlier symptoms.

The purpose of this study is the first to determine the results of performing the DBSCAN method when viewed from the silhouette in classifying JKN participation in the NTB Province, and the second aim to determine the results of the cluster formed using the DBSCAN method in classifying JKN participation in the NTB Province.

B. LITERATURE REVIEW

Data mining is known as Knowledge Discovery in Database (KDD). It is also defined as the process which includes extracting the interesting, interpretable and useful information from the raw data (Madni et al., 2017). In data mining, the term spatial data mining is known, which is defined as finding interesting and previously unknown but potentially useful patterns from large spatial datasets (Regin et al., 2021).

Clustering and classification are both fundamental tasks in Data Mining. Classification is used mostly as a supervised learning method, clustering for unsupervised learning (some clustering models are for both). The goal of clustering is descriptive, that of classification is predictive (Alelyani et al., 2018) (Dogan and Birant, 2021). Since the goal of clustering is to discover a new set of categories, the new groups are of interest in themselves, and their assessment is intrinsic. In classification tasks, however, an important part of the assessment is extrinsic, since the groups must reflect some reference set of classes (Koksalmis and Kabak, 2019). According to (Kameshwaran and Malarvizhi, 2014) (Wierzchoń and Kłopotek, 2018), clustering is grouping a set of data objects into multiple groups (clusters) so that objects in a group have high similarities, but differ from objects in other groups. Clustering is partitioning a set of data objects (observations) into subsets that can organize search results into groups and present the results in a concise and easily accessible way (Nagpal et al., 2013) (Wierzchoń and Kłopotek, 2018).

There are many clustering algorithms, one of which is Density-Based Clustering. Density Based Clustering can determine clusters based on irregular data shapes and can handle noise data effectively. Density-Based Spatial Clustering of Application with Noise (DBSCAN) is a development of density-based clustering techniques. DBSCAN is significantly more effective in finding clusters with arbitrary shapes, and can also find clusters with indeterminate shapes (Devi et al., 2015) (Zhang, 2019).

DBSCAN is a clustering method that builds an area based on density connected, meaning that every object that falls within the radius of the area must contain at least a minimum amount of data, so that noisy objects are not included in the cluster (Devi et al., 2015) (Zhang, 2019).

The elements in the clustering analysis process in the DBSCAN algorithm include (Zhang, 2019):

1. Epsilon
   Epsilon kinship from a profile or Eps-neighborhood from a profile Neps (p), is defined as:

$$N_{eps}(p) = \{q \in D \mid dist(p,q) \leq Eps\} \tag{1}$$

   D = analyzed database
   q = other profiles
   Eps = threshold value of the distance between profiles to be included in the same cluster.
2. Minimum Points MinPts is a threshold value that represents the minimum number of profiles in the Eps-neighborhood profile p in order to form a cluster (Zhang, 2019). After the grouping is done, the next step is to evaluate the results of the grouping using Silhouette Index cluster validation. The Silhoueutte index is calculated as confidence in the clustering process on an observation with a cluster that is said to be well-formed if the index value is close to 1 and the opposite condition if the index value is close to -1.

$$s_{(i)} = \frac{b_{(i)} - a_{(i)}}{\max(a_{(i)}, b_{(i)})} \tag{2}$$

   With
   $a_{(i)}$ = the average distance between i and all other observations
   $b_{(i)}$ = the average distance between i and observations in the nearest cluster
   Sillhouette values in the range -1 to 1.

C. RESEARCH METHOD

1. Data Source

The data used in this study was got from the official website of the National Population and Family Planning Agency (BKKBN) (https://www.bkkbn.go.id). The data comprises two variables, namely BPJS PBI and BPJS Non PBI. BPJS PBI is a participant in the Health Insurance Contribution help (PBI) which includes people with the poor and the underprivileged, as mandated by the SJSN Law whose contributions are paid by the government (Riyadi, 2018), while BPJS Non PBI are participants who are not classified as poor and poor people, who pay their dues individually or collectively to BPJS Health (Riyadi, 2018).

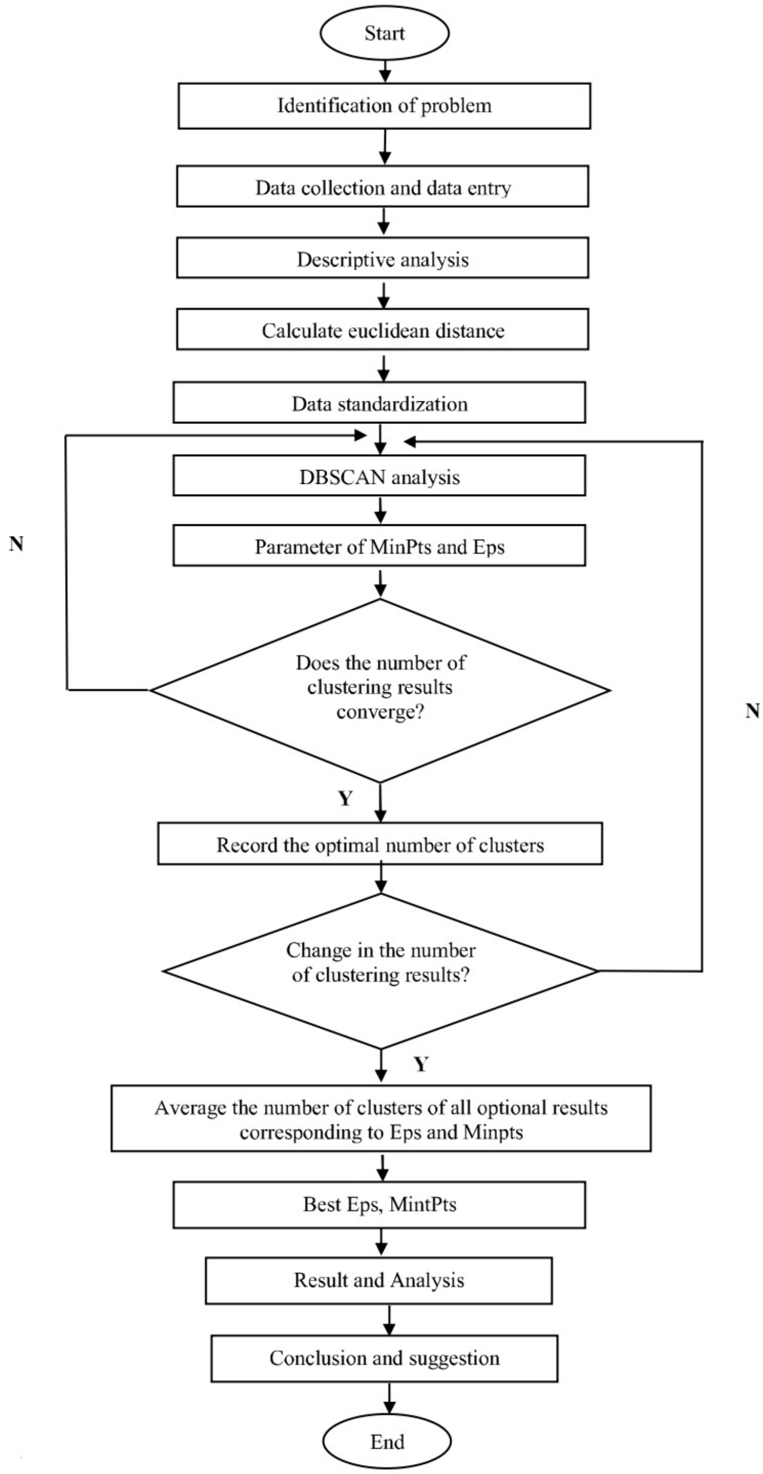

Figure 1. Research methodology flow chart

The analytical technique used in this research is descriptive analysis and DBSCAN using R software. The DBSCAN Algorithm, according to (Hou et al., 2016)(Song and Lee, 2018) is:

1. Initialize Minpts parameters, Eps.
2. Determine the starting point (p) at random.
3. Repeat steps 35 until all points are processed.
4. Calculate Eps or all distances of points that are density reachable regarding p.
5. If the point that meets Eps is more than Minpts, then the point p is the core point and a cluster is formed.

6. If p is a border point, and no point is density reachable regarding p, then the process is continued to another point.

D. RESULTS AND DISCUSSION

1. Descriptive Analysis

a. Percentage of JKN Participation

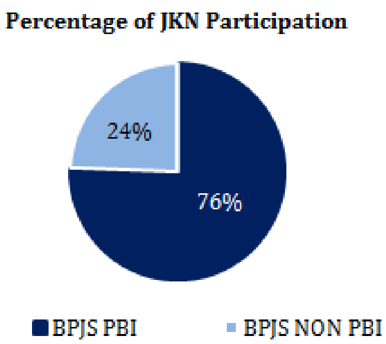

Figure 2. Percentage of JKN Participation

Figure 2 shows the percentage of participation in the National Health Insurance. BPJS PBI participants are the largest percentage, namely 76%, while BPJS Non PBI participants are 24%. According to (Andria and Kusnadi, 2017) there is a tendency for BPJS PBI participants to be poor and near poor in using public and private health facilities that use health insurance.

b. BPJS Participation Box plot

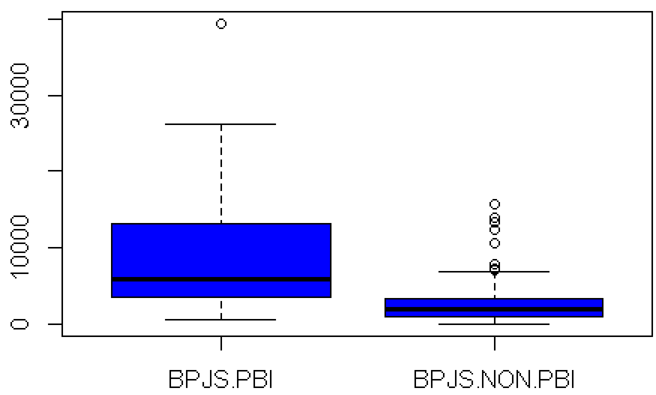

Figure 3. BPJS Participation Box plot

Figure 3 is a box plot of BPJS participation. Based on the box plot, we can see that there are outliers for each variable. BPJS PBI has one outlier, while BPJS Non PBI has more outlier data than BPJS PBI. Therefore, the DBSCAN method is very suitable to be used in the analysis of these data.

2. Forming a Distance Matrix

In determining the measure of distance commonly used in cluster analysis is the Euclidean distance. If P_1 and P_2 are 2 observations that are at a distance of Rn, then the Euclidean distance is presented in the equation below:

$$d_{Euclid}(P_1, P_2) = \sqrt{\sum_{i=1}^{n}(P_{1i} - P_{2i})^2} \tag{3}$$

The calculation is to determine the distance or similarity between one object and another or one sub-district with another. The results of the calculation of the similarity distance from one sub-district to another can be seen in the following matrix:

$$D = \begin{bmatrix} 0,0000 & 3990,16 & 1651.31 & 16302 & 4232,06 & \cdots & 15740 \\ 3990,16 & 0,0000 & 3740,72 & 13177 & 3191,53 & \cdots & 3964,9 \\ 1651.31 & 3740,72 & 0,0000 & 14988 & 2800,73 & \cdots & 5750.5 \\ 16302 & 13177 & 14988 & 0,0000 & 12189,1 & \cdots & 5069,5 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 15740 & 3964,97 & 5750,5 & 5069,5 & 6594,08 & \ddots & 0,0000 \end{bmatrix}$$

From this matrix, an example of the results of the Euclidean distance between Gerung and Kediri sub-districts is 3990.16, while the Euclidean distance between Kediri and Narmada sub-districts is 1651.31. Likewise, for interpreting other objects, the smaller the distance between the two objects, the more similar the characteristics of the two objects.

3. Density-Based Spatial Clustering of Application with Noise (DBSCAN) Analysis

The clustering process is carried out using the DBSCAN method. The clustering process in this study is:

a. Data Standardization

Standardization is done if the data used in a study has units that vary or are different. For this reason, it then carried the standardization process out by transforming the original data before further analysis.

b. Determining Epsilon and MinPts

The initial determination of the value of these parameters should get the best cluster. The determination of the Eps1, Eps2, and MinPts values influences the resulting cluster, which is determined by the Eps and MinPts values at points in a cluster. The point in question is that the k closest neighbors are roughly the same distance the noise point has the furthest distance from the k-nearest neighbors so that a plot of the distances is carried out sequentially at each point in the k-nearest neighbors. To get the optimal Eps and MinPts values, the line that is close to the ascending line is chosen and then cut vertically into the k-nearest neighbor plot. In this study, the value of $k = 2$. We will enter this value into the DBSCAN algorithm to determine the optimal number of clusters. The following is Figure 4 of the k-dist plot to determine the epsilon value

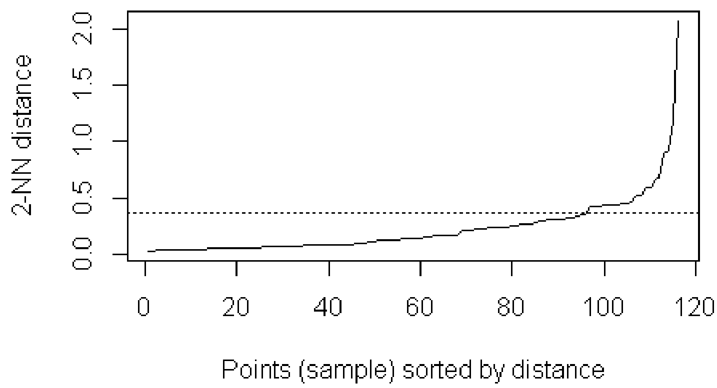

Figure 4. Plot of determining the value of epsilon

Figure 4 is a KNN plot, it is known that the optimal epsilon value is in the 0.37 point range by observing the knee of a curve or the elbow curve is the point where the curve looks bent, especially from high slopes to low slopes or in other directions. Used in optimization if the points are optimal for some decision, for example, when there is an increase from vertical to horizontal.

DOI: https://doi.org/10.30812/varian.v6i1.1886

The optimal selection of Epsilon and MinPts parameters is seen from the highest Silhouette value by using several experiments implemented in the R software. It showed experimental values of Epsilon and MinPts on BPJS Health Participation data in Table 1.

Table 1. DBSCAN Experiment

Epsilon	MinPts	Cluster	Noise	Silhouette Index
0.37	2	5	7	0.1894
0.45	2	4	6	0.1396
0.50	2	4	5	0.1336
0.37	3	3	13	0.2763
0.45	3	2	10	0.2586
0.50	3	2	9	0.2497

Based on the experimental results in Table 1, the epsilon values of 0.37 and MinPts 3 were chosen because they had the highest silhouette values, namely 0.2763.

c. **Forming a Cluster**

Previously, an analysis of the determination of epsilon and MinPts has been carried out and the determination of the optimal number of clusters by looking at the silhouette index value. It showed the distribution of the number of clusters in Figure 5 and Table 1, which is a cluster table from the analysis results got from the data processing using the DBSCAN method.

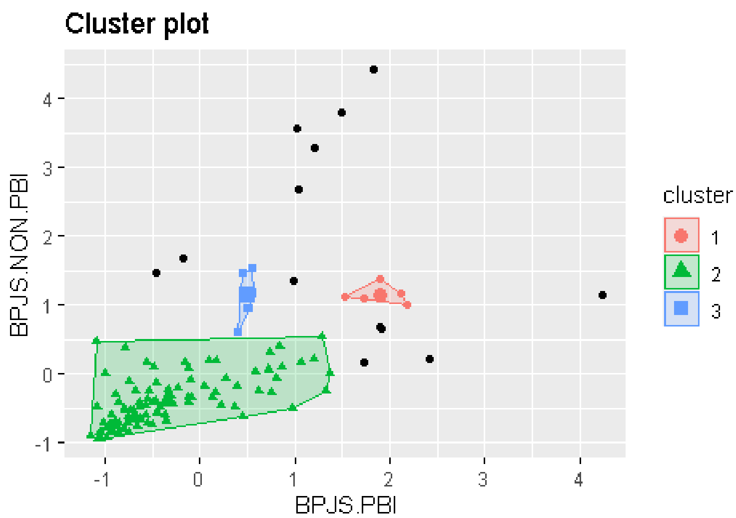

Figure 5. Cluster Plot

In Figure 5 there are 3 clusters generated. The first cluster is a red dot with 5 districts. Cluster 2 is the green dot with 93 sub-districts and cluster 3 is the blue dot with 5 sub-districts, while the black dot represents noise, which is 13 sub-districts. The division of clusters for sub-districts in the province of NTB based on BPJS Health participation is as follows.

DOI: https://doi.org/10.30812/varian.v6i1.1886

Table 2. Cluster Results

Cluster	Member
Noise	Kediri, Praya, Jonggat, Masbagik, Selong, Pringgabaya, Aikmel, Sumbawa, Ampenan, Mataram, Cakranegara, Selaparang, and Mpunda
Cluster1	Gerung, Narmada, Labuapi, Gunung Sari, and Sandubaya
Cluster2	Sekotong, Lingsar, Sheet, Batu Layar, Kuripan, Batukliang, Pujut, West Praya, East Praya, Janapria, Pringgarata, Kopang, Southwest Praya, North Batukliang, Keruak, Sakra, Terara, Sikur, Sukamulia, Sambelia, Montong Gading, Pringgasela , Suralaga, Wanasaba, Sembalun, Suela, Labuhan Haji, East Sakra, West Sakra, Jerowaru, Lunyuk Alas, Utan, Batu Lanteh, Moyo Hilir, Moyo Hulu, Ropang, Lape, Plampang, Empang, Alas Barat, Labuhan Badas, Labangka, Buer, Rhee, Unter Iwes, Moyo Utara, Maronge, Tarano, Lenang-guar, Orong Telu. Lantung, Kempo, Hu'u, Kilo, Woja, Concentrated, Manggalewa, Pajo, Monta, Bolo, Woha, Belo, Wawo, Sape, Wera, Donggo, Sanggar, Ambalawi, Langgudu, Lambu, Madapangga, Tambora, Soromandi, Parado, Lambitu, Palibelo, Jereweh, Seteluk, Sekongkang, Brang Rea, Poto Tano, Brang Ene, Maluk, Tanjung, Ganga, Kayangan, Bayan, and Pemenang,
Cluster3	Praya Tengah, Dompu, Taliwang, Sekarbela, and Raba

The results of clusters in this study show that there are different characteristics between clusters, the identification analysis is shown in Table 3 and Figure 6.

Table 3. Cluster Results

	PBI	Non PBI
Cluster1	22458.6	6240.4
Cluster2	6476.8	1712.4
Cluster3	12460.8	6198

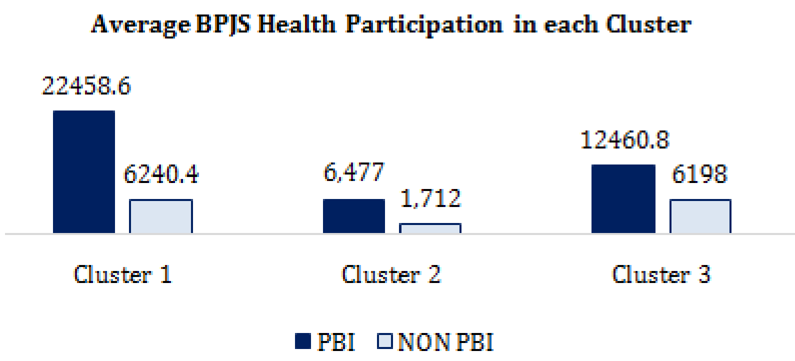

Figure 6. Average BPJS Health Participation in each Cluster

Figure 6 is a diagram that describes the characteristics of each variable. The interpretation of the characteristics of each cluster is as follows:

1. The highest JKN BPJS participants are in cluster 1 with 5 sub-districts. Cluster 1 has an average of 22,458 PBI participants and an average of 6420 Non-PBI participants. In cluster 1, the highest JKN BPJS participants are sub-districts in West Lombok Regency, so it can be said that in these sub-districts the distribution of BPJS participants is good.
2. The second highest cluster is cluster 3 with 5 sub-districts. Cluster 3 has an average of 12,460 BPJS PBI participants and 6198 Non PBI participants. Similar to cluster 1, in cluster 3 BPJS Health participants are quite high, so it can be said that they are good.
3. The lowest cluster is cluster 2 with 93 sub-districts. This cluster is the cluster with the lowest average JKN BPJS participants for PBI and Non PBI, with an average of 6476 and 1712, respectively. In contrast to clusters 1 and 3, BPJS Health participants in cluster 2 are still low. Seen 93 sub-districts with a low distribution of BPJS Health participants, especially in the dominating sub-district with the largest population in East Lombok Regency. This could be because of the unequal distribution of BPJS Health participants, and the lack of widespread participation in JKN, especially to BPJS Health who receive Contribution help from the government for the poor.

E. CONCLUSION AND SUGGESTION

The conclusion of this study results from the KNN plot that the optimal epsilon value is in the 0.37 point range through knee of a curve observation. And the experimental results for epsilon and MinPts values in determining the optimal cluster, selected epsilon values of 0.37 and MinPts 3 because they have the highest silhouette value of 0.2763. The highest JKN BPJS participants are in cluster 1 with 5 sub-districts with an average PBI participant of 22,458 participants and an average of 6420 Non PBI participants. The second highest cluster is cluster 3 with 5 sub-districts with an average of 12,460 BPJS PBI participants and 6198 Non PBI participants. While the lowest cluster is cluster 2 with 93 districts, 13 sub-districts are not included in any group because they are noise data.

This study uses DBSCAN Clustering with using a silhouette index to determine the optimal number of clusters. This can be an illustration for readers, both the community and institutions related to population and health, in increasing access to higher quality health services. We also hoped that this research can be developed, as well as become information so that policies are right on target.

REFERENCES

Aggarwal, C. C. (2017). An introduction to outlier analysis. In *Outlier analysis*, pages 1–34. Springer.

Alelyani, S., Tang, J., and Liu, H. (2018). Feature selection for clustering: A review. *Data Clustering*, pages 29–60.

Ali, A. and Masyfufah, L. (2021). Klasterisasi Pasien BPJS Dengan Metode K-Means Clustering Guna Menunjang Program Jaminan Kesehatan Nasional Di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo. *Jurnal Wiyata: Penelitian Sains dan Kesehatan*, 8(1):8–22.

Andria, F. and Kusnadi, N. (2017). Dampak kepesertaan bpjs bagi pekerja informal di bogor. *JIMFE (Jurnal Ilmiah Manajemen Fakultas Ekonomi)*, 3(1):1–15.

BPS NTB (2021). Profil Kesehatan NTB 2020. *In Profil Kesehatan NTB 2020, BPS Provinsi NTB*. https://ntb.bps.go.id/publication/2021/10/29/84e8b14e6e70ce610e55f109/profil-kesehatan-provinsi-nusa-tenggara-barat-2020.html.

Campello, R. J., Moulavi, D., Zimek, A., and Sander, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 10(1):1–51.

Devi, A. S., Putra, I. K. G. D., and Sukarsa, I. M. (2015). Implementasi metode clustering dbscan pada proses pengambilan keputusan. *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, pages 185–191.

Dogan, A. and Birant, D. (2021). Machine learning and data mining in manufacturing. *Expert Systems with Applications*, 166:114060.

Fisher, D., DeLine, R., Czerwinski, M., and Drucker, S. (2012). Interactions with big data analytics. *interactions*, 19(3):50–59.

Galán, S. F. (2019). Comparative evaluation of region query strategies for dbscan clustering. *Information Sciences*, 502:76–90.

Gebremeskel, G. B., Yi, C., He, Z., and Haile, D. (2016). Combined data mining techniques based patient data outlier detection for healthcare safety. *International Journal of Intelligent Computing and Cybernetics*.

Hendayanti, N. P. N., Putri, G. A. M. A., and Nurhidayati, M. (2018). Ketepatan klasifikasi penerima beasiswa stmik stikom bali dengan hybrid self organizing maps dan algoritma k-mean. *Jurnal Varian*, 2(1):1–7.

Herlinda, V., Darwis, D., and Dartono, D. (2021). Analisis clustering untuk recredesialing fasilitas kesehatan menggunakan metode fuzzy c-means. *Jurnal Teknologi Dan Sistem Informasi*, 2(2):94–99.

Hou, J., Gao, H., and Li, X. (2016). Dsets-dbscan: A parameter-free clustering algorithm. *IEEE Transactions on Image Processing*, 25(7):3182–3193.

Kameshwaran, K. and Malarvizhi, K. (2014). Survey on clustering techniques in data mining. *International Journal of Computer Science and Information Technologies*, 5(2):2272–2276.

Kha, N. H. and Anh, D. T. (2015). From cluster-based outlier detection to time series discord discovery. In *Trends and Applications in Knowledge Discovery and Data Mining*, pages 16–28. Springer.

Koksalmis, E. and Kabak, Ö. (2019). Deriving decision makers weights in group decision making: An overview of objective methods. *Information Fusion*, 49:146–160.

Madni, H. A., Anwar, Z., and Shah, M. A. (2017). Data mining techniques and applicationsa decade review. In *2017 23rd international conference on automation and computing (ICAC)*, pages 1–7. IEEE.

Malkomes, G., Kusner, M. J., Chen, W., Weinberger, K. Q., and Moseley, B. (2015). Fast distributed k-center clustering with outliers on massive data. *Advances in Neural Information Processing Systems*, 28.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing type i error and power in linear mixed models. *Journal of memory and language*, 94:305–315.

Nagpal, A., Jatain, A., and Gaur, D. (2013). Review based on data clustering algorithms. In *2013 IEEE conference on information & communication technologies*, pages 298–303. IEEE.

Pourbahrami, S., Balafar, M. A., Khanli, L. M., and Kakarash, Z. A. (2020). A survey of neighborhood construction algorithms for clustering and classifying data points. *Computer Science Review*, 38:100315.

Regin, R., Rajest, S. S., and Singh, B. (2021). Spatial data mining methods databases and statistics point of views. *Innovations in Information and Communication Technology Series*, pages 103–109.

Riyadi, E. D. A. (2018). Analisis tingkat penyediaan jaminan sosial untuk petugas k3l di lingkungan universitas padjadjaran. *Focus: Jurnal Pekerjaan Sosial*, 1(2):87–96.

Sadewo, T. A., Kusuma, P. D., and Setianingsih, C. (2021). Clustering pada data sentimen bpjs kesehatan menggunakan algoritma agglomerative hierarchical clustering (ahc) average linkage. *eProceedings of Engineering*, 8(5).

Saky, D. A. L., Jayanti, N. A., and Nurmayanti, W. P. (2020). Clustering Petani Berdasarkan Dampak Covid-19 Yang Terjadi Pada Sektor Pertanian. In *Seminar Nasional Official Statistics*, volume 2020, pages 160–164.

Song, H. and Lee, J.-G. (2018). Rp-dbscan: A superfast parallel dbscan algorithm based on random partitioning. In *Proceedings of the 2018 International Conference on Management of Data*, pages 1173–1187.

Wierzchoń, S. T. and Kłopotek, M. A. (2018). *Modern algorithms of cluster analysis*, volume 34. Springer.

Zhang, M. (2019). Use density-based spatial clustering of applications with noise (dbscan) algorithm to identify galaxy cluster members. In *IOP conference series: earth and environmental science*, volume 252, page 042033. IOP Publishing.

Zhao, Q., Shi, Y., Liu, Q., and Fränti, P. (2015). A grid-growing clustering algorithm for geo-spatial data. *Pattern Recognition Letters*, 53:77–84.

Zimek, A. and Filzmoser, P. (2018). There and back again: Outlier detection between statistical reasoning and data mining algorithms. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(6):e1280.