JITTER - Jurnal Ilmiah Teknologi dan Komputer Vol.6, No.2 August 2025

Clustering K-Means for Data Distribution of Treatment Classes, Provinces, and Facilities

Ihsanul Fikri[a1], **Evta Indra**[a2]
[a]Faculty of Technology and Computer Science, Universitas Prima Indonesia, Indonesia,
e-mail: [2]evtaindra@unprimdn.ac.id

Abstrak

BPJS Kesehatan adalah badan hukum yang menyelenggarakan program Jaminan Kesehatan Nasional dengan tujuan memberikan akses layanan kesehatan kepada seluruh rakyat Indonesia. Meskipun memberikan banyak manfaat, distribusi kepesertaan BPJS Kesehatan masih belum merata di seluruh wilayah Indonesia. Penelitian ini bertujuan untuk menganalisis sebaran peserta BPJS Kesehatan dengan menerapkan algoritma K-Means Clustering. Proses analisis melibatkan serangkaian proses preprocessing data, penerapan Elbow Method untuk menentukan jumlah cluster yang optimal, hingga evaluasi kualitas cluster menggunakan metode Davies-Bouildin Index (DBI). Hasil penelitian ini menunjukkan bahwa lima cluster adalah yang terbaik dengan DBI senilai 0.50 yang menunjukkan bahwa klasterisasi telah terbentuk dengan baik, jelas, dan terpisah.

Kata kunci: *K-Means Clustering, Davies-Bouildin Index, Metode Elbow, BPJS Kesehatan*

Abstract

BPJS Health is a legal entity that organizes the National Health Insurance program with the aim of providing access to health services to all Indonesian people. Despite providing many benefits, the distribution of BPJS Kesehatan membership is still uneven throughout Indonesia. This study aims to analyze the distribution of BPJS Kesehatan participants by applying the K-Means Clustering algorithm. The analysis process involves a series of data preprocessing processes, applying the Elbow Method to determine the optimal number of clusters, and evaluating cluster quality using the Davies-Bouildin Index (DBI) method. the results of this research show that five clusters are the best with a DBI of 0.50 which shows that the clustering has been well formed, clear and separated.

Keywords: *K-Means Clustering, Davies-Bouildin Index, Elbow Method, National Health Insurance Program*

1. Introduction

Every citizen is entitled to a prosperous life, physical and mental well-being, adequate housing, and a healthy living environment [1]. BPJS Health is a legal entity formed with the aim of organizing the National Health Insurance (JKN) program for all Indonesian people [2]. Therefore, the Indonesian government has established a National Health Insurance (JKN) program through the Health Social Security Administration Agency or BPJS Health (BPJS) with the aim of providing guarantees for all Indonesian people to get basic health needs [3]. In the last 10 years, the number of BPJS Health participants reached 267,784,196 people [4]. The existence of BPJS Health has a positive impact on the community, especially the underprivileged. Among them is by reducing public expenditure in accessing health services in Indonesia [5].

However, the pattern of uneven distribution of BPJS participants. Although the BPJS health program has reached 72% of the Indonesian population [6]. However, there is still an uneven distribution of participants in several regions. Therefore, an in-depth analysis is needed for clustering using the K-Means algorithm on the distribution pattern of participants. Clustering is the process of grouping similar objects into different groups, or partitioning a data set into several subsets, so that the data in each subset depends on a predetermined size or distance [7]. With the application of this method, BPJS Health and the government can formulate new policies in the equitable distribution of BPJS Health participants.

2. Research Method

2.1 Data Acquisition

The available data is sample data on BPJS Health membership for the period 2016-2022 which can be accessed through the National Health Insurance (JKN) data portal website. This dataset consists of 18 columns that cover various variables regarding BPJS Health membership.

2.2 Data Used

The data used are 8 variables that will be employed to cluster the participants' distribution pattern using the K-Means algorithm, namely Age, Gender, Participant Province, Participant Regency, Health Facility Province, Health Facility City, Participant Treatment Class and Participant Segmentation.

2.3 General Architecture

The General Architecture used is to describe the methodology in research that is used in a structured manner.

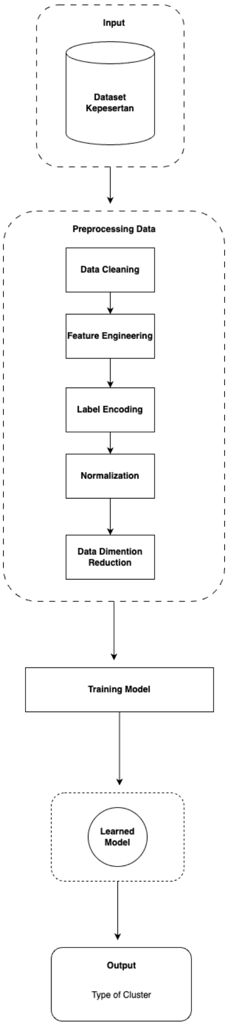


Figure 1. General Architecture

1. Data preprocessing is a step taken to process raw data into data that is suitable for use in the model training process. Data preprocessing is an important stage that serves as the main foundation to ensure validity in data analysis. This process includes a series of techniques and methods designed to improve the quality of imperfect raw data. These

steps include various aspects such as removing outliers, handling missing values, and refining the data to make it more consistent and clean, so that the resulting data is ready for use in further analysis [8]. The data preprocessing stage consists of data cleaning, feature engineering, label encoding, and data normalization.

a. Data cleaning, a process used to identify data that is inaccurate, incomplete, or non-reasonable, then improves its quality through the correction of detected errors and flaws [9]. There are 2 processes in it, namely Missing Value and Duplicate. Missing values may occur due to human errors during data processing, machine malfunctions, or respondents' refusal to answer specific questions, exit of study participants, unrelated data aggregation, and various other problems [10]. Some of the techniques commonly used to deal with missing values are deletion, imputation, and the use of built-in methods. Duplication can be defined as the same entity having two or more representations in a database [11]. A common way to deal with duplicate values is to delete any duplicate or identical data entries. If missing values are found, these values are changed to 'NONE' to indicate that the required information is not available. In this study, if duplicate data is found, the lines are removed to maintain the integrity and quality of the data.

b. Feature engineering, one of the key features of machine learning to extract useful patterns from data that will make it easier for the model to differentiate classes [12]. This process involves creating, modifying, and selecting features in the data to produce a better model.

c. Label encoding, a method that converts categorical data represented as text labels into numerical formats. it entails mapping each category or label in a categorical variable to a unique numerical value [13]. One of its commonly used applications is through the label encoder feature provided by the scikit-learn library. The stages for processing non-numerical data into numerical data. This is done because machine learning algorithms can only process data represented in numerical form. By converting data to numerical data, models can more easily recognize patterns and perform analysis.

Table 1. Application of Label Encoding

Before Label Encoding	After Label Encoding
Class I	0
Class II	1
Class III	2

Table 1 shows the changes and implementation of the label encoding process. Each category of text is converted to numerical values. Class I categories are changed to 0, Class II to 1, and Class III to 2.

d. Data normalization, the use of the Min-Max Scaler method is carried out so that machine learning algorithms can work better and avoid potential underfitting and overfitting.

Table 2. Examples of Data to Be Normalized

PSTV09	**PSTV10**
26	321
33	236
18	287
9	462
13	269

Table 2 shows two variables in the PSTV09 and PSTV10 columns that have different value ranges. Because of these range differences, normalization is required to ensure that both columns are on the same scale. The technique used to normalize the column is Min-Max Normalization. The first step in Min-Max normalization is to specify the minimum and maximum values of each column. In column PSTV09, the minimum value is 9 and the maximum value is 33, while in column PSTV10, the minimum value is 236 and the maximum is 462. The following step involves calculating the value of the normalization result using Equation 1 below:

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 1)$$

For the value of 26 in the first row in column PSTV09, the value of normalization results is calculated using equation 1 as follows:

$$x' = \frac{26 - 9}{33 - 9} = \frac{17}{24} = 0{,}7083$$

Furthermore, for the value of 321 or the first row in the PSTV10 column, the value of the normalization result is calculated with equation 1 as follows:

$$x' = \frac{321 - 236}{462 - 236} = \frac{85}{226} = 0{,}3761$$

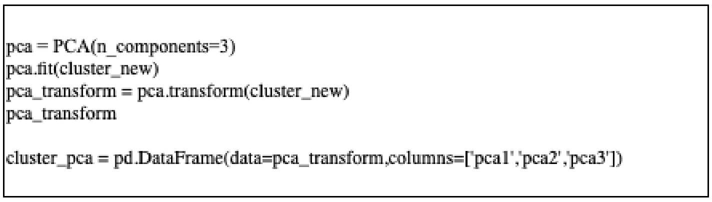
The above steps are repeated on each data in each column until the data normalization results are obtained as displayed in table 3 below:

Table 3. Results of the Implementation of Min-Max Normalization

PSTV09	**PSTV10**
0,7083	0,3761
1	0
0,375	0,2257
0	1
0,1667	0,146

Table 3 above shows a fairly long range of values, now all are in a more consistent data range, namely 0 and 1.

e. Data reduction, to simplify the complexity of the dataset into 8 variables, data reduction was carried out using Principal Component Analysis (PCA). The code snippets are as shown in Figure 2 below:



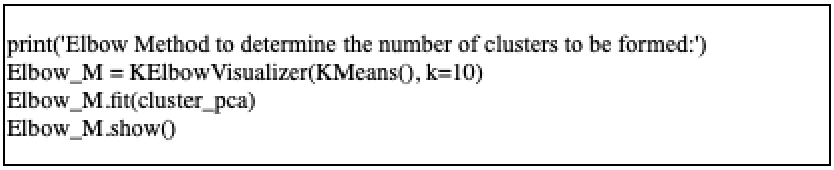
```
pca = PCA(n_components=3)
pca.fit(cluster_new)
pca_transform = pca.transform(cluster_new)
pca_transform

cluster_pca = pd.DataFrame(data=pca_transform,columns=['pca1','pca2','pca3'])
```

Figure 2. Data Reduction

Figure 2 above shows data that has been previously normalized and still has 8 variables reduced to 3 variables using the code pca = PCA (n_components=3). Once the fitting is done, it further transforms the dataset using the code pca_transform = pca.transform(cluster_new. It is then stored into a new dataset named cluster_pca using the code cluster_pca = pd. DataFrame(data=pca_transform, columns=['pca1', 'pca2', 'pca3']).

2. Training Process, at this stage, the K-Means algorithm will be used to classify the processed data. K-Means is an unsupervised learning algorithm that aims to group data into clusters based on the similarities between the data. Before applying the K-Means algorithm, an important step is to determine the optimal number of clusters using the elbow method. *The elbow* method is used to find the right number of clusters by measuring intra-cluster variation (the distance between data within the cluster) against the number of different clusters.
3. The Elbow method, a technique used to select the most suitable number of clusters by analyzing a graph that shows the relationship between the number of clusters and the distortion score value, which is the total distance between the data and the center of the cluster. The first step in the Elbow Method is to determine the range of k-values, which is the number of clusters to be tested, usually starting from 2 to a higher number, such as 10. The code snippet looks like Figure 3 below:



```
print('Elbow Method to determine the number of clusters to be formed:')
Elbow_M = KElbowVisualizer(KMeans(), k=10)
Elbow_M.fit(cluster_pca)
Elbow_M.show()
```

Figure 3. Elbow Method

Figure 3 above shows the process starting with the initialization *of the Elbow Visualizer* using the object Elbow_M = KElbowVisualizer(KMeans(), k=10) which sets the maximum limit of the cluster to be tested. Next, this visualization is applied to the dataset that has been reduced with PCA, i.e. cluster_pca using the Elbow_M.fit(cluster_pca) method.

Once the fitting process is complete, the result is displayed with Elbow_M.show() which generates an elbow graph. In the graph, the x-axis represents the number of clusters, while the y-axis represents the Within-Cluster Sum of Squares (WCSS) or the total distance between each data point and its cluster centroid.

4. K-Means Clustering, to group BPJS Health data, K-Means will divide data into several clusters based on the similarity between the data, according to the number of clusters that have been predetermined. The code snippet looks like the following Figure 4:

```
kmean=KMeans(n_clusters=5)
kmean.fit(cluster_pca)
labels =kmean.labels_
cluster_pca=pd.concat([cluster_pca, pd.DataFrame({'cluster':labels})], axis=1)
```

Figure 4. K-Means Clustering

Figure 4 above illustrates the grouping of data into five clusters with the code kmean = KMeans(n_clusters=5). Furthermore, the K-Means algorithm was applied to the dataset that had been reduced using PCA, namely cluster_pca using the .fit() method. Once the fitting is complete, the cluster labels for each data are obtained via labels = kmean.labels. The code shows the cluster label for each data. Then the training results are stored in a new dataset that still uses the name cluster_pca in which there is an addition of cluster label columns. This process is done using the code cluster_pca = pd.concat([cluster_pca, pd. DataFrame({'cluster': labels})], axis=1) that adds a cluster column to the cluster dataset_pca. Thus, each row of data now has a cluster label that facilitates further analysis.

5. The evaluation stage is carried out to evaluate the performance of the clustering model that has been produced using the Davies-Bouldin Index (DBI) method**.** To calculate the DBI value, it is done using the davies_bouldin_score function from the sklearn.metrics library. This function will calculate the DBI based on the clustering results generated by the K-Means algorithm, taking into account the resulting cluster labels and the grouped data as seen in Figure 5 below:

```
db_index = davies_bouldin_score(cluster_pca, labels)
```

Figure 5. Value of DBI

Figure 5 above shows that the DBI value is calculated using the davies_bouldin_score() function from the sklearn.metrics library. This function accepts two main parameters: datasets that have been reduced using PCA(cluster_pca) and cluster labels resulting from the K-Means algorithm (labels). The calculation process starts by calling davies_bouldin_score(cluster_pca, labels) and the results are stored in db_index variables.

3. Literature Study

Previous research has been researched by Azkar, et all with the title Clustering of Outpatients in Health Centers Using the K-Means Clustering Algorithm Method [14]. In his research at the North Aikmel Health Center in East Lombok Regency, patient data has not been used for decision-making to improve the quality of health services. This study grouped 1889 outpatient data based on gender, age, BPJS membership status, address, and disease diagnosis using the K-Means algorithm and the RapidMiner application. Evaluation with the Davies Bouldin Index (DBI) showed that two clusters were optimal: Cluster 1 with 1570 patients, dominated by women (68.4%), BPJS participants (52.2%), adults (56.2%), from Toya village (31.1%), with the most diagnoses of acute ISPA (14.2%); and Cluster 2 with 319 patients, dominated by women (64.3%), BPJS participants (63.3%), adults (59.9%), from Toya village (33.9%), with the most diagnoses of pregnancy examinations (9.1%). This information is expected to help the Puskesmas in making decisions or policies related to health programs.

In addition, a research conducted by Widyaningsih et all entitled The Application of the K-Means Clustering Method in Grouping the Number of BPJS Health JJKN/KIS Participants in Cirebon Regency [15]. The study aims to group BPJS Health or JKN/KIS participants in Cirebon Regency based on data from the West Java Open Dataset consisting of 412 participants using the K-Means Clustering method with RapidMiner tools. The results of this study grouped BPJS Health participants into 3 clusters. The results of the grouping were assessed with the Davies Bouildin Index (DBI) for optimization and a value of K=3 was obtained as the optimal value with DBI = 0.164.

In addition, another study has also been conducted by Khairani et all with the title Clustering of BPJS Participants Based on Medical Records Using the K-Means Algorithm [16]. Based on medical records, the study seeks to assist Dumai Hospital group BPJS participants in determining the condition, gender, and dominant BPJS group. The K-Means Clustering algorithm was used to process medical record data of inpatient BPJS PBI patients from January to December 2022. The results created three clusters: Cluster 0 was dominated by diseases with code O342 (Maternal care due to uterine scar from previous surgery), which affected women aged 21–40 years; Cluster 1 was dominated by diseases with code E119 (Diabetes mellitus without complications), which affected women aged 41–60 years; and Cluster 2 was dominated by diseases with code J180 (Bronchopneumonia, unspecified organism), which affected females aged 41–60 years.

4. Result and Discussion

The implementation process generally uses the Pyhton programming language, Google Colab and Jupiter Notebook.

4.1. Model Evaluation

After the K-Means Clustering model is successfully formed with 5 (five) clusters, the next step is to conduct an evaluation to assess the quality of the model that has been built. The method used is the Davies-Bouldin Index (DBI) as an evaluation metric, which is one of the common indices to measure cluster validity. The DBI method evaluates how well the clusters are formed by measuring the ratio between the average distance between clusters and the distance between the center of the cluster to the points in the cluster. The lower the DBI value, the better the quality of the clusters produced, as it shows that the clusters are well separated from each other. DBI is calculated using Equation 2 below:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i \neq j}\ (R_{i,j}) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 2)$$

After the K-Means Clustering model is successfully formed with 5 (five) clusters, the next step is to conduct an evaluation to assess the quality of the model that has been built. The method used is the Davies-Bouldin Index (DBI) as an evaluation metric, which is one of the common indices to measure cluster validity. The DBI method evaluates how well the clusters are formed by measuring the ratio between the average distance between clusters and the distance between the center of the cluster to the points in the cluster. The lower the DBI value, the better the quality of the clusters produced, as it shows that the clusters are well separated from each other. DBI is calculated using Equation 2 below:

4.2. Treatment Class Distribution

The treatment class is divided into 3, namely treatment class 1, treatment class 2, and treatment class 3. To gain further understanding of the distribution of treatment classes in each cluster, visualizations were carried out in the form of bar charts. This visualization helps show how each class of care is distributed within each of the clusters that have been formed like Figure 6 below:

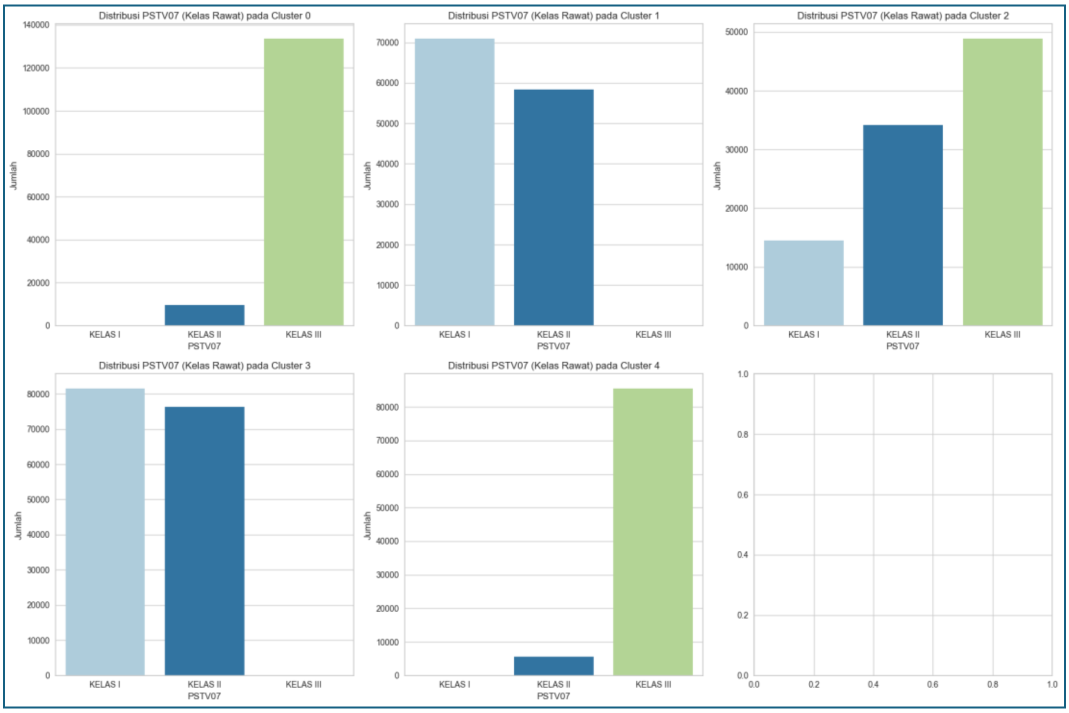


Figure 6. Distribution of Treatment Classes

From Figure 6 above, it shows a graph of variation in the number of participants spread across different treatment classes and clusters. In cluster 0, inpatient class III dominated with the highest number, while class I and class II had much less. Meanwhile, cluster 1 and cluster 3 show a fairly balanced variation in number between class I and class II. Meanwhile, in clusters 2 and 4, it can be seen that class II again dominates.

4.3. Distribution of Participants' Provinces of Residence

BPJS Health participants are spread across all provinces in Indonesia, for this reason, visualizations are given to see the distribution of participants between provinces in each cluster.

The visualization is displayed in the form of a heatmap that shows the number of participants by province of origin for each cluster as shown in Figure 7 below:

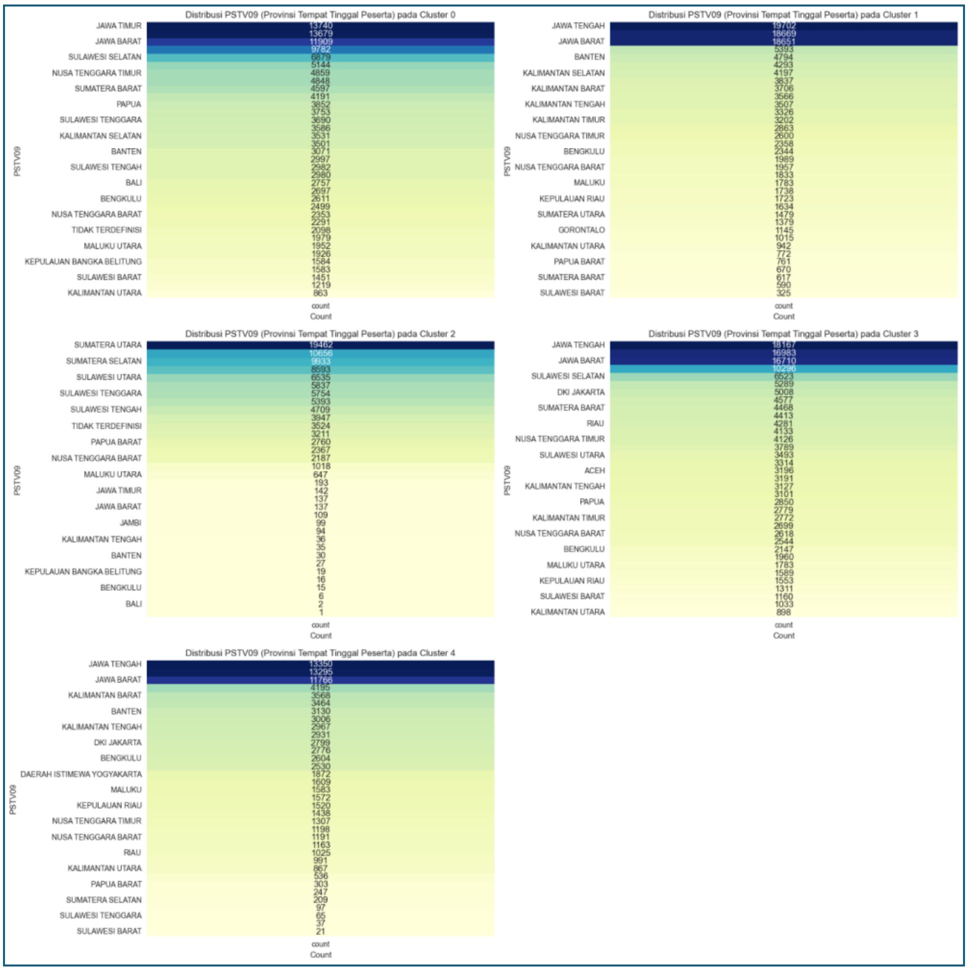


Figure 7. Distribution of Participants' Provinces of Residence

In Figure 7 above, it can be seen that in cluster 0, the dominance of participants comes from the provinces of East Java, West Java, and South Sulawesi, while provinces such as North Kalimantan, West Sulawesi, and the Bangka Belitung Islands have a relatively smaller number of participants. Meanwhile, in cluster 1, it can be seen that the participants are dominated by the provinces of Central Java, West Java, and Banten, and there are several provinces with a smaller number of participants, such as North Kalimantan and West Sulawesi.

Meanwhile, cluster 2 is dominated by North Sumatra, South Sumatra, and North Sulawesi while provinces such as Bali and the Bangka Belitung Islands have a smaller number of participants. Cluster 3 shows a similar distribution, with Central Java, West Java, and South Sulawesi as the provinces with the highest number of participants. Some other provinces have a more even distribution, with North Kalimantan and the Riau Islands having fewer participants.

Finally, cluster 4 has the most participants from Central Java, West Java, and West Kalimantan, while provinces such as West Sulawesi and South Sumatra have fewer participants.

4.4. Provincial Distribution of Registered Health Facilities

Furthermore, a visualization of the distribution of health facility registration places by BPJS Health participants was made in the form of *a heatmap* as shown in Figure 8 below:

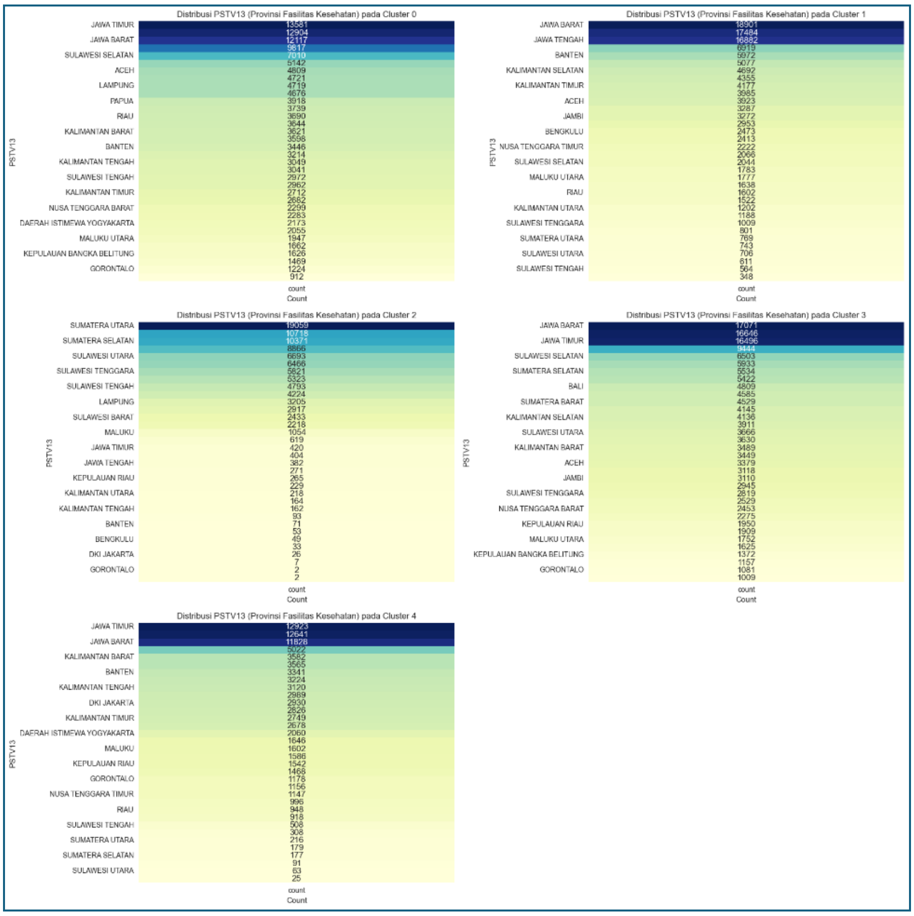


Figure 8. Distribution of Provincial Health Facilities

In Figure 8 above, it can be seen that in cluster 0, it can be seen that the provinces of East Java, West Java, and South Sulawesi have the highest concentration of health facilities registered by participants, shown by darker colors on the *heatmap*. In contrast, provinces such as Gorontalo and the Bangka Belitung Islands had a smaller number of health facilities registered by participants, as seen from the lighter colors.

In cluster 1, it can be seen that participants in the provinces of West Java, Central Java, and Banten dominate compared to participants from health facilities in other provinces. Meanwhile, there are fewer participants who registered for health facilities in the provinces of North Sumatra, North Sulawesi, and Central Sulawesi.

In cluster 2, it can be seen that the dominant participants are registered in health facilities in the provinces of North Sumatra, South Sumatra, and North Sulawesi. On the other hand, the provinces of Bengkulu, DKI Jakarta, and Gorontalo do not have enough participants registered at their health facilities.

In Cluster 3, it again shows the provinces of West Java, East Java, and South Sulawesi with the largest number and the Bangka Belitung and Gorontalo Islands with the smallest number.

In Cluster 4, it was dominated by participants who registered at health facilities located in East Java, West Java, and West Kalimantan, while provinces such as Gorontalo, East Nusa Tenggara, and Riau showed a smaller number.

5. Conclusion

The conclusions of this study include:

1. The K-Means Clustering model with 5 clusters shows quite good performance in grouping data. The results of the model evaluation showed that the clusters formed had a clear structure and separated the data well.

2. The Davies-Bouldin Index (DBI) model is 0.50. This value indicates that the clusters that form have a relatively clear separation from each other. Lower DBI indicates that the clusters are more well separated, indicating good clustering quality.
3. Further analysis of the distribution of treatment classes and the province of residence and facilities of participants showed that there was a high concentration of BPJS Health participants in several provinces. This shows that some provinces have a larger number of participants compared to others.
4. Data shows that there are provinces with low BPJS Health membership. This underlines the need for additional focus to increase awareness and access to BPJS Health in these provinces. Increased efforts in promotion and outreach in low-participation areas can help in expanding coverage and improving access to health services for more participants.

References

[1] BAPPENAS RI, “Negara Republik Indonesia Tahun 1945,” *Warga dan Negara*, vol. 1945, pp. 1–166, 1945.

[2] A. Rahim, S. A. Fajriah, S. Diniah, V. I. Sabilah, and S. Suryadi, “Implementasi Pengawasan Bagi Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan Ditinjau dari Hukum Administrasi Negara,” *JIIP - J. Ilm. Ilmu Pendidik.*, vol. 6, no. 8, pp. 5795–5800, 2023, doi: 10.54371/jiip.v6i8.2573.

[3] A. F. Diba and G. Oka Warmana, “Sosialisasi Bpjs Kesehatan Pada Pegawai Security Untuk Mengetahui Pentingnya Jaminan Kesehatan Di Ciputra Surabaya,” *J. Pengabdi. Kpd. Masy.*, vol. 2, no. 3, pp. 198–202, 2022, [Online]. Available: https://jurnalfkip.samawa-university.ac.id/karya_jpm/index

[4] BPJS Kesehatan, “BPJS Kesehatan.” 2024. [Online]. Available: https://bpjs-kesehatan.go.id/#/

[5] E. A. Djamhari *et al.*, “Defisit Jaminan Kesehatan Nasional (JKN): Mengapa dan Bagaimana Mengatasinya ?,” *Prakarsa*, pp. 67–68, 2020, [Online]. Available: https://repository.theprakarsa.org/media/302060-defisit-jaminan-kesehatan-nasional-jkn-m-4c0ac9c6.pdf

[6] A. Ahdiat, “Persentase Penduduk yang Tak Punya Jaminan Kesehatan Berdasarkan Provinsi (2023),” databoks. Accessed: Jul. 15, 2024. [Online]. Available: https://databoks.katadata.co.id/datapublish/2023/12/21/jaminan-kesehatan-belum-merata-ini-kondisinya-di-tingkat-provinsi

[7] M. G. H. Omran, A. P. Engelbrecht, and A. Salman, “An overview of clustering methods,” *Intell. Data Anal.*, vol. 11, no. 6, pp. 583–605, 2007, doi: 10.3233/ida-2007-11602.

[8] R. Khoirina and F. Akhmad, “Sikap Sopan Santun - 7116-21857-2-Pb,” no. 2014, pp. 250–255, 2021.

[9] F. A. Rahman, R. Kassim, Z. Baharum, H. A. M. Noor, and N. A. Haris, “Data Cleaning in Knowledge Discovery Database-Data Mining (KDD-DM),” *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6s3, pp. 2196–2199, 2019, doi: 10.35940/ijeat.f1100986s319.

[10] P. Karkare, “Handling Missing Values in Data,” *Medium.Com*, no. October, pp. 750–756, 2019, [Online]. Available: https://medium.com/x8-the-ai-community/handling-missing-values-in-data-54e1dc77e24f

[11] G. V. Dhivyabharathi and S. Kumaresan, “A survey on duplicate record detection in real world data,” *ICACCS 2016 - 3rd Int. Conf. Adv. Comput. Commun. Syst. Bringing to Table, Futur. Technol. from Arround Globe*, pp. 1–5, 2016, doi: 10.1109/ICACCS.2016.7586397.

[12] E. Dariato, “Analisis dan Perancangan Feature Engineering Pada SQL Execution Plan Aplikasi Apache Spark,” *Arcitech J. Comput. Sci. Artif. Intell.*, vol. 1, no. 2, p. 83, 2021, doi: 10.29240/arcitech.v1i2.4120.

[13] C. Herdian, A. Kamila, and I. G. Agung Musa Budidarma, “Studi Kasus Feature Engineering Untuk Data Teks: Perbandingan Label Encoding dan One-Hot Encoding Pada Metode Linear Regresi,” *Technol. J. Ilm.*, vol. 15, no. 1, p. 93, 2024, doi: 10.31602/tji.v15i1.13457.

[14] Azkar and Kusrini, “KLIK: Kajian Ilmiah Informatika dan Komputer Klasterisasi Pasien Rawat Jalan di Puskesmas dengan Menggunakan Metode Algoritma Clustering K-

Means," *Media Online*, vol. 4, no. 5, pp. 2628–2636, 2024, doi: 10.30865/klik.v4i5.1832.

[15] T. A. WIDYANINGSIH and D. Pratama, "Penerapan Metode K-Means Clustering Dalam Mengelompokkan Jumlah Peserta Bpjs Kesehatan Jkn/Kis Di Kabupaten Cirebon," *E-Link J. Tek. Elektro dan Inform.*, vol. 18, no. 1, p. 17, 2023, doi: 10.30587/e-link.v18i1.5330.

[16] A. F. Khairani, A. Nazir, T. Darmizal, Y. Vitriani, and Y. Yusra, "Klasterisasi Peserta BPJS Berdasarkan Rekam Medis Menggunakan Algoritma K-Means," *J. Comput. Syst. Informatics*, vol. 4, no. 3, pp. 625–631, 2023, doi: 10.47065/josyc.v4i3.3442.