Contents lists available at IOCS

Science Midwifery

journal homepage: www.midwifery.iocspublisher.org

Cluster analysis of province case mix index on Indonesian national health security (BPJS Kesehatan) system

Raditya Prabaswara[1], Muhammad Ikhsan Kalla[2], Penbi Opmel Siregar[3], Muhammad Mizmara Alan Falihin[4], Benjamin Saut Parulian Simanjuntak[5], Samuel Sormin[6]

[1,2,4,5,6]Actuarial Deputy, BPJS Kesehatan, Jakarta, Indonesia

[3]Department of Actuary, Faculty of Mathematics and Natural Science, Institut Teknologi Sumatera, Lampung, Indonesia

ARTICLE INFO

Article history:

Received May 24, 2024
Revised May 28, 2024
Accepted Jun 7, 2024

Keywords:

Case Mix Index
Cluster Analysis
Clustering
K-Means Clustering
Hierarchical

ABSTRACT

Disease case is a diverse and complex problem which faced by health insurance institutions. BPJS Kesehatan, Indonesian national health security, has undoubtedly faced this problem, especially because of the socioeconomic gap between 34 provinces in Indonesia. To describe the diverse and complex conditions of resource needs for all hospital patients, this paper will calculate the Case Mix Index (CMI) for each province and conduct a cluster analysis to group provinces based on the CMI and total health facility visits similarity. Once the CMI for each province is identified, the cluster analysis is executed by using K-Means and Hierarchical clustering method to compare each result. The first step of cluster analysis is to identify the optimal number of clusters. In this article, several K-value selection techniques is used to find out the optimal number of clusters. By using K-value selection methods, the optimal number of clusters is two province clusters. The first cluster, namely Cluster 1, consists of four provinces which are DKI Jakarta, Jawa Barat, Jawa Tengah, and Jawa Timur. The second cluster, denoted as Cluster 2, consists of the rest provinces which are not included in Cluster 1. Although the optimal number of clusters are identified, this paper also adjust the cluster analysis result to provide comparison with the current INACBG regionalization. The result of this paper can be utilized as a recommendation for INA-CBGS tariff regionalization since using CMI as one of the clustering variables could depict the diverse condition in 34 provinces in Indonesia.

Corresponding Author:

Raditya Prabaswara,
Actuarial Deputy,
BPJS Kesehatan,
JL Letjen Suprapto Cempaka Putih, Jakarta, 10510, Indonesia,
Email: rprabaswara@gmail.com

INTRODUCTION

Indonesia is a developing country with around 262 million in population over 34 provinces with high variety of socioeconomic and health infrastructure conditions which could be depicted from the socioeconomic index and health infrastructure index which ranged between 40.66 to 75.16 and 25.47 to 77.53 respectively (Agustina et al., 2019). It is necessary to create a universal health

coverage (UHC) scheme that could accommodate these diverse needs and conditions. In response, by 2014, Indonesia launched a UHC programme called the National Health Insurance System (Jaminan Kesehatan Nasional/JKN) with Badan Penyelenggara Jaminan Sosial Kesehatan (BPJS Kesehatan) as the organizer (Agustina et al., 2019). Based on Undang-Undang Republik Indonesia No. 24 2011, BPJS Kesehatan is a government agency for health insurance that organize the JKN program. including managing membership, collecting premiums, administering contracts with providers and paying providers. BPJS Kesehatan has authority for the payment of health facilities based on the standard tariffs set by the government (Setiawan et al., 2023). Payments to primary care providers are through capitation, and to hospital providers through DRG episodes of service payments (INA-CBG) (Mahendradhata et al., 2017).

Indonesia is engaged in the process of ensuring effective decentralization to district level as through decentralization, it is expected that government can improve health service quality (Gunawan & Aungsuroch, 2015). The Indonesian health system consist of public and private facilities, from the provider and the financing services (Himawan, 2024). The public system works within decentralized government system in Indonesia, with central, provincial and district government responsibilities while the private facilities are networks of health facilities or financing service managed by not-for-profit and charitable organizations, and for-profit providers (Mahendradhata et al., 2017).

Based on Regulation of the Minister of Health of the Republic of Indonesia No. 52 2016 Indonesian-Case Based Groups (INA-CBG) tariff is the measurement of claim payment by BPJS Kesehatan to advanced referral health facility (Triyudawati et al., 2023), such as hospital and clinic, which formed from diagnosis related group (DRG) and procedure (Roesbiantoro et al., 2022). Currently, INA-CBG tariff is divided into five regionals which consist of provinces. In recent years, most of the private-owned hospital's patients are JKN members so that 80%-90% of private-owned hospital's revenue is depending from INA-CBG payment (Nurwahyuni & Setiawan, 2020). Aside from ensuring an adequate health funding, local governments play a key role in the operation of various components that support the implementation of the JKN. They also need to map the availability of the necessary skill-mix, equipment and other resources at different health facilities by providing the regionalization of referral nodes (Mahendradhata et al., 2017). Unfortunately, Ramadhan et al (2022) stated that there is a significant difference between the real hospital cost and INACBG tariff with the hospital tariff 7%-66% higher than INACBG tariff in some disease. Therefore, there is a need for INACBG tariff adjustment to address this issue.

To describe these complex conditions conditions, Case Mix Index is applicable as it is comprised the disease types, treatment methods, length of stay, cost, etc. Case Mix Index (CMI) is a value assigned to a DRG of patients in a healthcare. It is widely used by hospitals as an element to assess clinical complexity and difficulty of treatment. Moreover, CMI can also used to determine the allocation of healthcare resource for patients (Chang & Zhang, 2019). A health facility CMI is calculated from the sum of relative weights per each facility's DRG divided by the number of patients (Mendez et al., 2014).

In the recent years, there are plenty research that focused on the implementation of CMI and new regionalization recommendation with clustering approach. Aldino et al (2021) used K-Means algorithm for clustering planting area based on planting feasibility. This research involves the Euclidean Distance calculation for each region point and finally divided planting region into two clusters. Research by Abdullah et al (2022) used K-Means clustering method to group provinces based on the risk of the COVID-19 in Indonesia. In this research, Elbow, Silhouette, and Gap Statistic method was used to determine the optimal number of clusters. The result shown that there are three clusters of provinces or two clusters if DKI Jakarta is not included in the provinces list. Delamater et al. (2013) employ K-Means and Ward's clustering algorithm to group hospitals which create an objective, replicable, and sustainable method to group hospital in Michigan based on the proportion of patient visits in home areal units. Han et al (2018) used CMI to assess the

resource allocation of all cases in a hospital. This research examined the relationship between the CMI and nursing workload with Spearman Correlation Analysis. A research on hospital resource allocation by Wang & Gao (2017) used Case Mix Index to incorporate patient condition information as the heterogeneity of patients conditions is an important factor to analyse the efficiency of the resource allocation in a hospital. By using CMI, Xie et al (2022) developed a new hospital payment and management system, namely Big Data Diagnosis and Intervention Packet.

This study calculated outpatient and inpatient CMI for each province and developed a new regionalization system for INACBG tariff by using K-Means clustering method with outpatient and inpatient CMI as one of the clustering variables. Through this study, the new regionalization system will involve the clinical complexity and difficulty of disease treated in each province.

RESEARCH METHOD

In this study, Case Mix Index for 34 provinces in Indonesia is calculated by using 2017 to 2021 advance level health facilities utilization data from BPJS Kesehatan which include disease case and cost. To calculate the CMI, firstly, the Medicare Severity-Diagnosis Related Group (MS-DRG) weight are is calculated which in this research used INACBG code to describe the diagnosis. Next, the weights are summed up by province group and dividing the total by the number of discharges. After the CMI per provinces are determined, K-Means clustering is conducted. This study used outpatient and inpatient CMI as one of the clustering variables to describe the clinical complexity for each province. Moreover, this research also used Gross Regional Domestic Product (GRDP) to describe the economic condition. To support the result of K-Means clustering, this research used Silhouette and Elbow method to find the optimal number of clusters. Furthermore, the second and third most optimal number of clusters also collected to provide broader clustering alternatives.

Cluster Analysis with K-means Method

K-means clustering separate data space into K non-overlapping clusters. These clusters are represented by their centroids or the mean of every point inside the cluster. Let $D = \{x_1, \dots, x_n\}$ is the data. K-means is an objective function which depends on the distance between data points to the cluster centroids.

$$\min_{\{m_k\}, 1 \le k \le K} \sum_{k=1}^{K} \sum_{x \in C_k} \pi_x dist(x, m_k) \qquad (1)$$

With:

π_x = weight of x
n_k = Number of data objects assigned to cluster C_k
$m_k = \sum_{x \in C_k} \frac{\pi_x x}{n_k}$ is the centroid of C_k
K = Number of clusters.

On Equation 1, "dist" function calculate the distance between object x and centroid m_k (Wu, 2012). In this paper, stats package from R is used to conduct K-means clustering.

Optimal Number of Cluster

As mentioned earlier, one of the drawbacks of using K-Means method for cluster analysis is the K-Value selection. Therefore, to outcome this problem, there are several methods for selecting K-Value for K-Means clustering such as Elbow and Silhouette method. Although elbow method has simpler algorithm than the latter, it has some drawbacks in using Elbow method to select K-Value. Schubert (2023) stated that elbow method lack of theoretic support in specify the number of clusters from elbow plot since there is no measurement of angle resulting the difference between human interpretation. Consequently, instead of using Elbow method to select K-Value,

this research used two methods which are Silhouette method and a method proposed by Pham et al which used evaluation function.

Pham et al (2005) purposed a K-Value selection method using evaluation function $f(k)$ which defined by Equation 2 and Equation 3:

$$f(K) = \begin{cases} 1 \; if \; K = 1 \\ \frac{S_K}{\alpha_K S_{K-1}} \; if \; S_{K-1} \neq 0, \forall K > 1 \\ 1 \; if \; S_{K-1} = 0, \forall K > 1 \end{cases} \quad (2)$$

$$\alpha_K = \begin{cases} 1 - \frac{3}{4N_d} \; if \; K = 2 \; and \; N_d > 1 \\ \alpha_K + \frac{1-\alpha_{K-1}}{6} \; if \; K > 2 \; and \; N_d > 1 \end{cases} \quad (3)$$

With:

S_K = Sum of the cluster distortions
K = Number of clusters
N_d = Number of dimensions
α_K = Weight factor.

Every K with $f(K) < 0.85$ are the recommended K-Value and if there is no single K with $f(K) < 0.85$, then $K = 1$ is selected. In this paper, R package namely kselection is used to conduct this method.

Silhouette method was firstly introduced by Rousseeuw in 1987 (Rousseeuw, 1987). It is a method that uses average dissimilarity of object i to all other objects of cluster A $(a(i))$ and average dissimilarity of i to all objects of other cluster C $(d(i, C))$. After calculating $d(i, C)$ for all clusters $C \neq A$, select the smallest $d(i, C)$ and denote it by $b(i)$. The silhouette value is calculated with Equation 4 as folows:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}. \quad (4)$$

The silhouette value ranging from -1 to 1. The closer the value to 1, the closer the relationship between the object and the cluster (Yuan & Yang, 2019). In this research, the factoextra package from R is used to implement silhouette method.

RESULTS AND DISCUSSIONS

Before continue conduct cluster analysis, this research collected Gross Regional Domestic Product (GRDP) growth from 2017 to 2021 data from Badan Pusat Statistik.

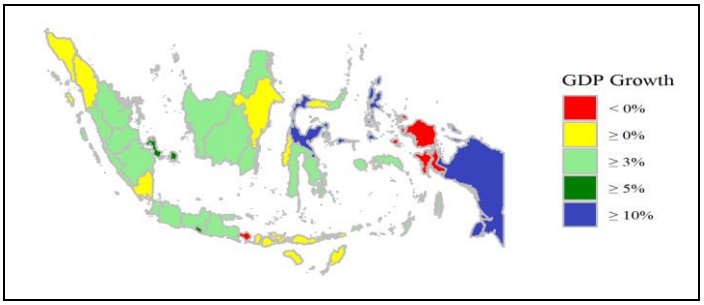

Figure 1. GRDP growth from 2017 to 2021 for each provinces

This research divided province's GRDP growth into five groups which are $< 0\%$, $\geq 0\%$, $\geq 3\%$, $\geq 5\%$, and $\geq 10\%$. Province with the highest GRDP growth is Maluku Utara (16.79%) while Bali has the lowest GRDP growth (-2.46%). There is a high variability in Indonesia GRDP growth since it has 3.62 standard deviation.

Case Mix Index

Inpatient and outpatient Case Mix Index for each province has been calculated by using INA-CBG weight. This research separate CMI into two groups which are < 1 and ≥ 1.

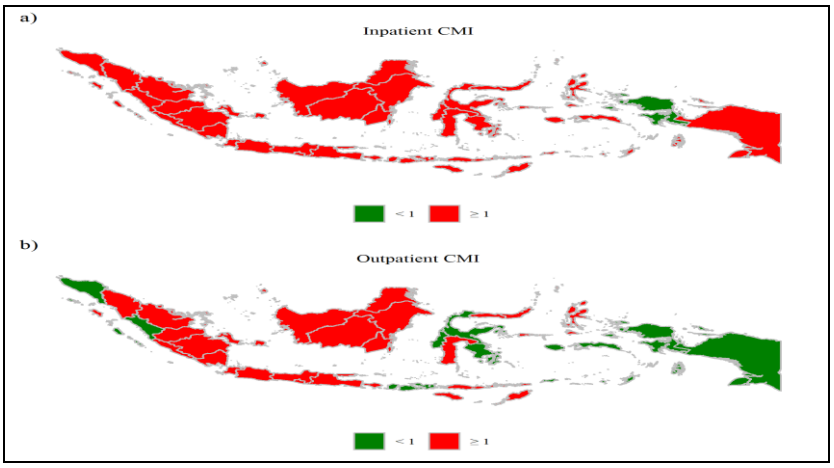

Figure 2. a) Outpatient CMI for each Province, b) Inpatient CMI for each Province

Based on Figure 2, Yogyakarta has the highest outpatient CMI and inpatient CMI with 1.33 and 1.50 respectively, while the lowest is Sulawesi Barat and Papua Barat for outpatient and inpatient CMI. The average outpatient and inpatient CMI in Indonesia are 1.08 and 1.20 respectively.

K-Value Selection

Evaluation Function

Based on the method proposed by Pham et al which use evaluation function $f(K)$ where K is the number of clusters, the recommended value of K is when $f(K)$ lower than 0.85. In this research, $K = \{1, \dots, 15\}$ and the value of $f(K)$ is shown on Table 1.

Table 1. Result of evaluation function $f(K)$ for each k-value

K	$f(K)$
1	1
2	0.252891
3	0.543464
4	0.774515
5	0.942323
6	1.074244
7	1.087912
8	0.305413
9	3.780725
10	1.041074
11	1.043356
12	1.038723
13	1.030498
14	1.01815
15	1.021853

It is shown in Table 1 that the most optimal number of clusters is 2 clusters since when $K = 2$, the value of $f(K) = 0.25$ which is the lowest $f(K)$ value. Furthermore, we also could take another K-value with $f(K) < 0.85$ such as 3, 4, and 8.

Silhouette Method

Silhouette method and evaluation function produced the same optimal K-value which is two clusters. However, there is slight difference between the other recommended result of the evaluation function and silhouette method.

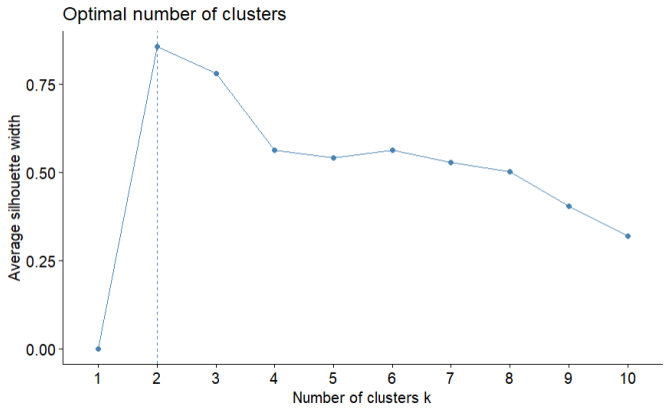

Figure 3. Silhouette method plot and the result of K-value selection

As stated before, the closer the silhouette value to 1, the closer the relationship between the object and its cluster. Therefore, As shown in Figure 3, the other recommended K-value are 3,4,5, and 6 as these numbers are the K-value which has the closest silhouette value to 1 besides $K = 2$.

K-means Clustering

The optimal K-means clustering result has provided in this research in accordance to the K-value selection result mentioned earlier which is $K = 2$. Therefore, provinces in Indonesia are divided into 2 clusters as shown in the Figure 4.

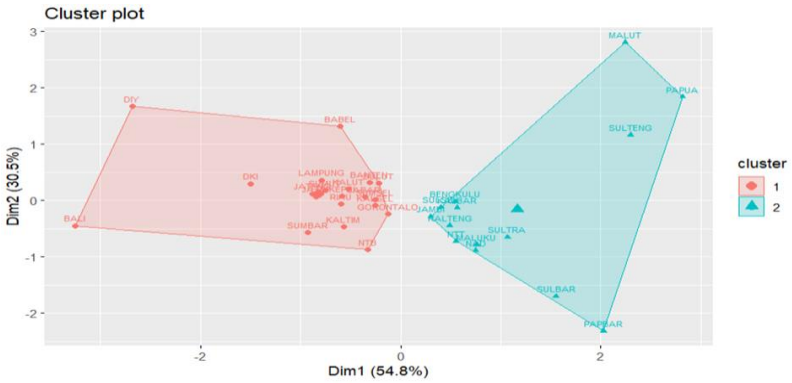

Figure 4. Optimal k-means clustering result

Since this paper used three variables or dimensions, R conduct Principal Component Analysis in fviz_cluster() command of factoextra package to visualize the clustering result in two dimension.

Furthermore, this research also provided an alternate clustering result by adjusting the K-value into 5 clusters as shown in Figure 5 This adjustment is done to compare the result of this paper and the current INACBG regionalization which is 5 regions.

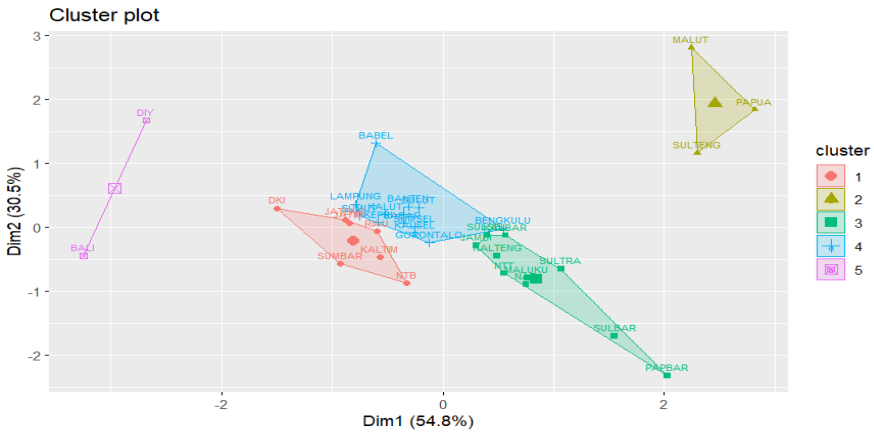

Figure 5. Optimal k-means clustering result

Discussion

Minister of Health Regulation No. 27 of 2014 stated that INACBG regionalization is based on regional Consumer Price Index (CPI). Therefore, there is a huge difference between the current regionalization and the result of this research as this research used three variables which are outpatient CMI, inpatient CMI, and GDP growth.

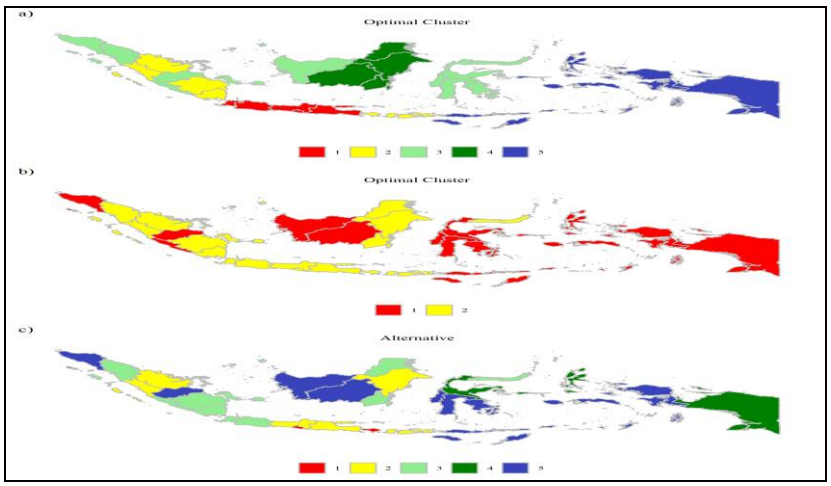

Figure 6. a) Map Plot of Optimal K-means Cluster Analysis, b) Map Plot of Adjusted K-Means Cluster Analysis c) Map Plot of Current Regionalization

CONCLUSION

This paper has calculated the Case Mix Index by using 2017 to 2021 disease case and cost of treatment data from BPJS Kesehatan and this paper also add 2017 to 2020 GDP growth data from Badan Pusat Statistik as one of the clustering variables. K-value selection techniques has done to determine the optimal number of clusters which produce $K = 2$ based on its CMI and GDP growth. K-means clustering method is used to conduct cluster analysis. Moreover, this paper also provided two results, the optimal result which based on the K-value selection and the adjusted result to provide comparison with current INACBG regionalization. There is a very significant difference

between the result of this paper and the current regionalization as both of them used different variable where the result of this paper used outpatient and inpatient CMI while the current regionalization use CPI.

This research could be used as a recommendation for INACBG regionalization system as it has more complete variable by using Case Mix Index as it comprised the clinical complexity and the resource allocation in health facilities. Moreover, by using GDP growth, it also included the economic growth condition in each province into INACBG regionalization system. For further research, several decision-making techniques such as Analytical Hierarchy Process (AHP) or Analytical Network Process (ANP) could be use in selecting clustering variables to provide more reliable result.

References

Abdullah, D., Susilo, S., Ahmar, A. S., Rusli, R., & Hidayat, R. (2022). The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data. *Quality and Quantity*, *56*(3). https://doi.org/10.1007/s11135-021-01176-w

Agustina, R., Dartanto, T., Sitompul, R., Susiloretni, K. A., Suparmi, Achadi, E. L., Taher, A., Wirawan, F., Sungkar, S., Sudarmono, P., Shankar, A. H., Thabrany, H., Susiloretni, K. A., Soewondo, P., Ahmad, S. A., Kurniawan, M., Hidayat, B., Pardede, D., Mundiharno, … Khusun, H. (2019). Universal health coverage in Indonesia: concept, progress, and challenges. In *The Lancet* (Vol. 393, Issue 10166). https://doi.org/10.1016/S0140-6736(18)31647-7

Aldino, A. A., Darwis, D., Prastowo, A. T., & Sujana, C. (2021). Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency. *Journal of Physics: Conference Series*, *1751*(1). https://doi.org/10.1088/1742-6596/1751/1/012038

Chang, J., & Zhang, L. (2019). Case Mix Index weighted multi-objective optimization of inpatient bed allocation in general hospital. *Journal of Combinatorial Optimization*, *37*(1). https://doi.org/10.1007/s10878-017-0204-3

Delamater, P. L., Shortridge, A. M., & Messina, J. P. (2013). Regional health care planning: A methodology to cluster facilities using community utilization patterns. *BMC Health Services Research*, *13*(1). https://doi.org/10.1186/1472-6963-13-333

Gunawan, J., & Aungsuroch, Y. (2015). Indonesia health care system and Asean economic community. *International Journal of Research in Medical Sciences*. https://doi.org/10.18203/2320-6012.ijrms20150231

Han, B., Chen, X., & Li, Q. (2018). Application of case mix index in the allocation of nursing human resources. *Journal of Nursing Management*, *26*(6). https://doi.org/10.1111/jonm.12593

Himawan, N. S. S. (2024). Literature Review: Analysis of Hospital Readiness in Facing Standard Inpatient Class (KRIS). *Pharmacology, Medical Reports, Orthopedic, and Illness Details*, *3*(1), 7–19. https://doi.org/10.55047/comorbid.v3i1.1116

Mahendradhata, Y., Trisnantoro, L., Listyadewi, S., Soewondo, P., Marthias, T., Harimurti, P., & Prawira, J. (2017). The Republic of Indonesia health system review. *Health Systems in Transition*, *7*(1).

Mendez, C. M., Harrington, D. W., Christenson, P., & Spellberg, B. (2014). Impact of hospital variables on case mix index as a marker of disease severity. *Population Health Management*, *17*(1). https://doi.org/10.1089/pop.2013.0002

Nurwahyuni, A., & Setiawan, E. (2020). Kinerja Rumah Sakit Swasta dengan Pembayaran INA-CBGs di Era Jaminan Kesehatan Nasional: Casemix, Casemix Index, Hospital Base Rate. *Jurnal Ekonomi Kesehatan Indonesia*, *4*(2). https://doi.org/10.7454/eki.v4i2.3822

Pham, D. T., Dimov, S. S., & Nguyen, C. D. (2005). Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *219*(1). https://doi.org/10.1243/095440605X8298

Ramadhan, L., Aritonang, M., & Anggriani, Y. (2022). Analisis Perbedaan Tarif Rumah Sakit dan Tarif INA-CBGs Pelayanan Rawat Jalan di RSUD Pasar Rebo Jakarta. *Journal of Islamic Pharmacy*, *6*(2). https://doi.org/10.18860/jip.v6i2.12147

Roesbiantoro, A., Setianto, B., Adriansyah, A. A., Asih, A. Y. P., Setiyowati, E., Bistara, D. N., & Sa'adah, N. (2022). Description Of Characteristics, Diagnosis And Financing Of BPJS Patients In ENT Poly Health Service Facility Level 2. *Medical Technology and Public Health Journal*, *6*(2), 169–176.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis.

Journal of Computational and Applied Mathematics, *20*(C). https://doi.org/10.1016/0377-0427(87)90125-7

Schubert, E. (2023). Stop using the elbow criterion for k-means and how to choose the number of clusters instead. *ACM SIGKDD Explorations Newsletter*, *25*(1). https://doi.org/10.1145/3606274.3606278

Setiawan, E., Witjaksana, B., & TJendani, H. T. (2023). Analysis of Risk Management in Building Workers of SMAN 5 Brawijaya Building Kediri. *International Journal on Advanced Technology, Engineering, and Information System*, *2*(4), 334–346.

Triyudawati, W., Kustriyani, A., & Prasasti, A. (2023). Analysis of Hospital and INA-CBG'S Rates of JKN Outpatient in Genteng Regional Public Hospital in 2021. *Proceedings of International Pharmacy Ulul Albab Conference and Seminar (PLANAR)*, *3*, 22–27.

Wang, D., & Gao, Q. (2017). Efficiency assessment and resource allocation for hospitals by data envelopment analysis. *2017 3rd International Conference on Information Management, ICIM 2017*. https://doi.org/10.1109/INFOMAN.2017.7950394

Wu, J. (2012). Advances in K-means Clustering: a data mining thinking. In *Springer Theses: recognizing outstanding Ph.D. Research*.

Xie, H., Cui, X., Ying, X., Hu, X., Xuan, J., & Xu, S. (2022). Development of a novel hospital payment system – Big data diagnosis & intervention Packet. *Health Policy OPEN*, *3*. https://doi.org/10.1016/j.hpopen.2022.100066

Yuan, C., & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. *J*, *2*(2). https://doi.org/10.3390/j2020016