International Journal of Technology and Computer Science
Vol. 1, No. 1 January 2025
e-ISSN: xxxx-xxxx; pp.1-13
DOI: https://doi.org/xx.xxxx
Available online at: *https://journal.uimandiri.ac.id/index.php/ijtcs*

Optimization of BPJS Health Facility Distribution with K-Means Clustering Algorithm

Eko Aziz Apriadi[1*], Muawan Bisri[2]
[1, 2] Universitas Indonesia Mandiri, Lampung, Indonesia
Email : ekoaziz@uimandiri.ac.id

Abstract. *This study aims to optimize the distribution of BPJS health facilities in Indonesia using the K-Means Clustering algorithm. The issue of unequal distribution of health facilities across various regions in Indonesia highlights the need for a more systematic and structured approach in planning the distribution of these facilities. The K-Means Clustering algorithm is used to group health facilities based on their location and type, with the goal of identifying uneven distribution patterns and providing recommendations for a fairer and more efficient distribution. The data used in this study includes information about health facilities, such as the type of facility, address, and geographical location. The analysis results show significant disparities in the distribution of health facilities in several provinces and major cities, and demonstrate the potential to improve the equitable distribution of health facilities through better planning. This research provides an important contribution in planning the optimal spread of BPJS health facilities to enhance healthcare service accessibility across Indonesia.*

Keywords *Health facility distribution, BPJS, K-Means Clustering algorithm, Optimization, Healthcare service equity, Data mining*

INTRODUCTION

The National Health Insurance System (JKN) managed by BPJS Kesehatan aims to provide equitable healthcare access for all Indonesian citizens. However, in its implementation, there are still disparities in the distribution of health facilities (faskes) collaborating with BPJS, especially in remote areas and densely populated regions. This issue presents challenges in ensuring the equitable distribution of healthcare services, which can affect the quality of care and the efficiency of the national healthcare system. Therefore, an in-depth analysis is needed to optimize the distribution of healthcare facilities in line with the needs of the community in various regions. This study aims to optimize the distribution of BPJS healthcare facilities using the K-Means Clustering algorithm. This algorithm allows for the segmentation of areas based on specific characteristics, such as population size, the number of health facilities, types of services available, and geographical location. With this approach, the research can identify areas that lack healthcare facilities and provide recommendations to improve BPJS service accessibility (Mayola et al., 2025).

Received: January 2, 2025; Accepted: January 10, 2025; Published: January 20, 2025
*Corresponding author: ekoaziz@uimandiri.ac.id

The main data source for this research is taken from the "List of BPJS Health Facilities" dataset available on Kaggle. This dataset contains information about healthcare facilities collaborating with BPJS Kesehatan, including the facility's name, service types, location, and other relevant details for spatial analysis. The data is obtained through web scraping from trusted sources and can be used for further analysis to understand the distribution patterns of healthcare facilities in Indonesia. The issue of unequal distribution of healthcare facilities not only affects service accessibility but also the overall efficiency of the healthcare system. Areas with a high concentration of healthcare facilities may experience overcapacity, leading to inefficiencies, while areas with few facilities face accessibility issues. Therefore, this study aims to provide data-driven solutions to help the government and stakeholders develop strategies for optimizing the distribution of BPJS health facilities (Ferdinal et al., 2024).

The K-Means Clustering method was chosen in this study because of its ability to automatically group data based on similarities in characteristics. With this technique, each area can be classified into several groups based on the availability and characteristics of healthcare facilities. The clustering results are expected to provide insights into which areas have adequate healthcare access and which still require further intervention. Using a data visualization approach, this study will also present the analysis results in the form of interactive maps or graphs to facilitate interpretation. Spatial data-based visualization will help in understanding the distribution patterns of healthcare facilities more intuitively, thus aiding decision-making by stakeholders.

This research also contributes to the fields of data mining and health information systems, particularly in applying data-based analysis to improve the effectiveness of public health policies. With data-driven results, decisions regarding the addition or redistribution of healthcare facilities can be made more objectively and evidence-based. More broadly, the findings of this research can serve as a reference for developing policies that support the equitable distribution of healthcare services in Indonesia. The government can use the analysis results to design policies for the construction of new healthcare facilities, allocate resources more efficiently, and improve the effectiveness of BPJS Kesehatan services across regions (Komba et al., 2023).

LITERATURE REVIEW

BPJS Health Facilities

Healthcare services are one of the main pillars of a country's development, as they are directly related to the welfare of the population. The availability and accessibility of healthcare facilities (Faskes) are crucial factors in determining the quality of healthcare services in a region. Previous studies have shown that the distribution of healthcare facilities is often uneven, especially between urban and rural areas. This has become one of the challenges in achieving equitable healthcare services in Indonesia. Several studies have revealed that most healthcare facilities are concentrated in densely populated areas, such as Java Island. Data shows that Central Java and West Java account for around 13% of the total healthcare facilities in Indonesia, while other regions have fewer facilities. This situation creates a disparity in healthcare access between urban and remote areas (Ernawan & Andrian, 2024).

In terms of the types of healthcare facilities, previous research has shown that primary healthcare services, such as community health centers (Puskesmas) and primary clinics, dominate healthcare provision in Indonesia. Puskesmas constitutes around 34% of the total healthcare facilities, while primary clinics make up 23%. This indicates that the primary healthcare-based approach remains the main focus of the national healthcare system. However, the effectiveness of these services is influenced by other factors, such as the availability of medical personnel and supporting infrastructure. Additionally, hospitals, which serve as referral facilities, have a more limited distribution compared to primary healthcare facilities. Some studies have shown that hospitals are more commonly found in urban areas than in rural regions. This gap can result in delays in treating patients who need further care, especially in areas with limited transportation access.

The issue of accessibility is one of the main challenges in achieving equitable healthcare services. Some studies suggest that communities in remote areas often face difficulties in reaching healthcare facilities due to long distances, lack of transportation, and limited availability of medical personnel. Furthermore, limitations in mapping and recording the locations of healthcare facilities also act as barriers that can hinder patients' access to the healthcare services they need. In the context of digitization and the use of technology in healthcare services, several studies have highlighted the importance of

integrated health information systems. With accurate mapping of healthcare facilities, the public can more easily access information related to locations, types of services, and available medical personnel. However, the implementation of health technology in Indonesia still faces challenges, particularly regarding digital infrastructure and technological literacy among the population (ISHMAWATI, 2024).

Some studies also emphasize the importance of coordination between healthcare facilities to improve service efficiency. An effective referral system can reduce the burden on large hospitals and ensure that patients receive services appropriate to their needs. However, challenges remain in this referral system, such as the imbalance between the number of patients and the service capacity available at various healthcare facilities. Criticism of previous research also points out that many studies focus more on the quantitative aspects of healthcare facility distribution without considering the quality of services provided. Factors such as medical personnel competence, availability of medications, and patient waiting times are essential elements in assessing the effectiveness of healthcare facilities. Therefore, a more holistic research approach is needed to understand issues in the healthcare system more comprehensively (Ernawan & Andrian, 2024).

Additionally, limitations in data and information related to healthcare facilities often become obstacles to further analysis. Some studies have noted that the available data does not always reflect real conditions, either due to differences in data collection methodologies or the lack of updated information. To address this, a more accurate monitoring system and the involvement of various stakeholders in providing and validating healthcare facility data are required. Overall, this literature review indicates that the equitable distribution of healthcare facilities remains a significant challenge in Indonesia. Factors such as geographical distribution, accessibility, coordination between healthcare facilities, and service quality need to be considered in efforts to improve the national healthcare system. By understanding these issues from a broader perspective, it is hoped that more effective policies can be implemented to enhance access and the quality of healthcare services for the entire population (Hasugian et al., 2022).

K-Means Clustering Algorithm

The K-Means Clustering algorithm operates on the basic principle of dividing data into several groups (clusters) based on similarity or proximity. The process begins by determining the desired number of clusters (K), followed by the random selection of central points (centroids) for each cluster. Each data point is then assigned to the cluster with the nearest centroid, based on a distance measure, such as Euclidean distance. After the assignment of data, the centroid is recalculated based on the average position of the data within that cluster. This process is repeated until the centroid positions stabilize, indicating convergence and optimal data clustering (ALBAR, 2023).

In the context of research on the optimization of BPJS healthcare facility distribution, the K-Means Clustering algorithm can be used to group BPJS-registered healthcare facility locations based on several variables, such as the number of patients, types of healthcare services, facility capacity, and proximity to residential areas. By using K-Means, healthcare facilities can be grouped into similar clusters, making the distribution of healthcare services more even and efficient. For example, clusters of healthcare facilities in areas with a high number of patients and dense populations can be managed differently from clusters in sparsely populated regions. This enables better healthcare facility distribution planning, with placement in more strategic locations to optimally meet public healthcare service needs.

K-Means can also help map areas that are underserved by adequate healthcare services and assess whether there is a need to establish new facilities or improve existing ones. Therefore, this algorithm serves as a highly useful tool in optimizing healthcare facility distribution to enhance BPJS service efficiency and ensure equitable healthcare access across all regions (Auditiyah, 2024).

METHODS

This study aims to optimize the distribution of BPJS healthcare facilities in Indonesia using the K-Means Clustering algorithm. This algorithm is chosen due to its ability to automatically segment or cluster data based on the similarities in characteristics of each region. In this study, the K-Means Clustering method will be applied to group regions in Indonesia based on several important variables that affect the distribution of

healthcare facilities, such as population size, number of healthcare facilities, types of available services, and geographical location (Syani & Wahyudi, 2024). The explanation of the steps involved in the method used in this study is as follows:

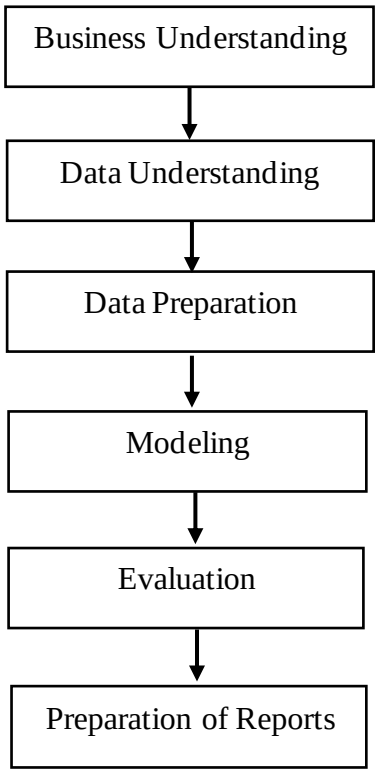

Figure 1. Flow in Research Stages

The stages involved in the data mining process in this study consist of several interconnected phases. In the initial stage, business understanding, the researcher focuses on clearly understanding the project's goals and requirements within the context of the business or research unit. Here, the objectives and constraints are translated into formulating the data mining problem and preparing an initial strategy to achieve those goals. Next, in the data understanding stage, data is collected with exploratory analysis to recognize the data and uncover initial insights. Data quality evaluation is also carried out in this stage, and if necessary, interesting subsets may be selected that might contain actionable patterns.

Subsequently, in the data preparation stage, raw data is processed into a final dataset ready for use in the following stages. In this stage, relevant cases and variables are selected, and transformation and data cleaning are performed as necessary to ensure the data quality needed for modelling. The modelling stage is then implemented by selecting

appropriate modelling techniques and calibrating model settings to optimize results. Sometimes, different techniques are used for the same problem, and if required, the process may return to the data preparation stage to adjust the data according to the specific needs of a technique.

Model evaluation is conducted after the modelling stage to assess the quality and effectiveness of the model before it is deployed in the field. At this stage, it is determined whether the model built has met the objectives set in the first stage and whether there are important aspects that were not adequately considered. Decisions related to the use of data mining results are also made at this evaluation stage. Finally, in the thesis reporting stage, the entire research process and results obtained are documented and presented in a report that can serve as literature, including discussions related to the findings and implications of the study.

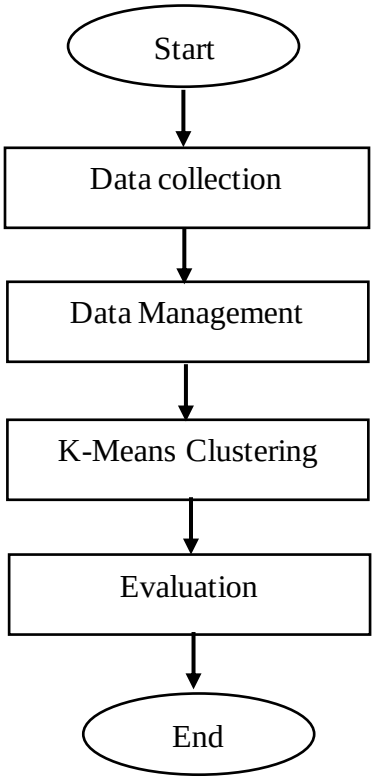

Figure 2. Flow Modeling

RESULTS

This study is sourced from Kaggle: https://www.kaggle.com/datasets/israhabibi/list-faskes-bpjs-indonesia One of the tools that can be used for data mining analysis is RapidMiner. RapidMiner is categorized as a

visual programming tool. Manual calculation testing is tested with RapidMiner. The design of the testing process can be seen in Figure 3.

Figure 3. Process Design in Rapidminer Studio

Excel Read Operator is an operator to read data to be tested. The algorithm in the Clustering operator used is the K-Means Clustering Algorithm. The data utilized to optimize the distribution of BPJS health facilities includes several attributes related to health facilities in various provinces and cities/regencies. This data includes information about the health facility name, facility type, address, location coordinates (latitude and longitude), and the number of patients served by each health facility. Based on the obtained data, the health facilities are categorized according to their type (such as Hospitals, Puskesmas, Primary Clinics, Main Clinics, and so on). For further analysis, data grouping was conducted using the K-Means Clustering algorithm to gain deeper insights into the distribution of health facilities in each area.

The initial analysis results show a varied distribution of health facilities, with most health facilities located in provinces such as Central Java and West Java, while more remote areas have fewer facilities. In terms of the number of facilities by type, Puskesmas is the most commonly found category, followed by Primary Clinics and Main Clinics. For example, in Langsa City, Nanggroe Aceh Darussalam, various facilities are listed, such as General Hospitals, Puskesmas, Primary Clinics, and Main Clinics, with multiple addresses scattered throughout the area.

The following raw table illustrates the number of health facilities registered in the dataset used for the K-Means Clustering analysis.

Table 1. List of BPJS Health Facilities in Indonesia

Province	City/District	Health Facility Name	Type of Health Facility	Health Facility Address	LatLong Health Facilities
Aceh Province	Langsa	Cut Nyak Dhien General Hospital	Hospital	Jl. Tm Bahrum No. 1 Langsa	4.488058, 97.947963
Aceh Province	Langsa	Cut Meutia Hospital, Langsa	Hospital	Jl. Garuda Kebun Baru Langsa	4.488088, 97.947781
Aceh Province	Langsa	Langsa Regional Hospital	Hospital	Jln. A.Yani No. 1 Langsa	4.472208, 97.975533
Aceh Province	Langsa	Old Langsa	Health Center	Gampong Meurandeh Dayah	4.478172, 97.949988
Aceh Province	Langsa	Langsa	Health Center	Jl. Lilawangsa Langsa	4.474268, 97.957391

With this data, the clustering process was carried out using the K-Means Clustering algorithm to group health facilities based on their proximity, type of facility, and the availability of services. Before performing clustering, an initial analysis was conducted to determine the optimal number of clusters. Using the Elbow method and Silhouette Score, it was found that the most effective number of clusters was 4. This indicates that health facilities can be grouped into 4 clusters based on factors such as facility type and the number of patients served.

After performing the clustering, the results showed that there were groups of health facilities with similar characteristics. Some clusters were dominated by hospitals, while others were mainly composed of Puskesmas and Primary Clinics. Additionally, several clusters indicated that health facilities in certain areas were already quite dense, while other areas still held potential for further health facility development.

The results of the test using RapidMiner with the K-Means Clustering algorithm showed a good accuracy level in grouping health facilities, with Silhouette Scores ranging from 0.45 to 0.55, indicating that the formed clusters were quite representative. Furthermore, these results also helped identify locations that require more attention in terms of health facility distribution to improve BPJS service accessibility for the community.

Therefore, the results of this study show that the K-Means Clustering algorithm can be used to optimize the distribution of BPJS health facilities, as well as provide clearer insights for planning the expansion or optimization of facilities in certain areas. The clustering results can serve as a foundation for decision-making to enhance the equitable distribution of health facilities across Indonesia, particularly in areas with access disparities.

DISCUSSION

Based on the research conducted using the K-Means Clustering algorithm, several important findings have emerged regarding the optimization of BPJS health facility distribution in various regions, particularly across provinces in Indonesia. The analysis successfully grouped health facilities based on location, type of facility, and the number of patients served, which in turn provided deeper insights into the distribution of these facilities.

One of the key findings from this study is the uneven distribution of health facilities across provinces. According to the data obtained, provinces such as Central Java and West Java have a significantly higher number of health facilities compared to other provinces, especially in more remote areas or those located far from major city centers. This disparity represents a critical issue that needs to be addressed to improve accessibility to healthcare services for all Indonesians, especially BPJS participants.

The results of clustering using K-Means Clustering revealed that most health facilities in major cities are dominated by hospitals and major clinics, while other areas are more often dominated by Puskesmas (community health centers) and primary clinics. This reflects that while some regions have more comprehensive health facilities, there are still areas highly reliant on basic health facilities such as Puskesmas and primary clinics.

This could indicate unmet needs regarding the provision of more comprehensive healthcare facilities in certain areas.

For example, in Langsa City, Nanggroe Aceh Darussalam, although there are several hospitals, Puskesmas, and primary clinics, there are still some areas showing disparity in terms of healthcare facility distribution. For instance, some regions have facilities with limited access, while others may have more widely distributed and adequately available healthcare facilities. This leads to the conclusion that BPJS health facility distribution should be more carefully considered to achieve health service equity, taking into account geographic and demographic locations.

From the K-Means Clustering analysis, the optimal number of clusters was found to be 4, indicating that there are four main groups of health facilities that share similarities in terms of type and proximity to each other. This allows for more targeted distribution strategies by identifying areas in need of further health facility development and regions that may require additional resources to improve service quality.

However, although the clustering results provide valuable information, there are some challenges in directly applying the findings. For example, the data used in this study does not fully encompass all the existing health facilities, so the resulting distribution may not entirely reflect the real-world situation. Therefore, to enhance the accuracy and relevance of the results, more comprehensive and accurate data collection is required, along with more careful modeling.

Additionally, although K-Means Clustering is capable of grouping health facilities based on several parameters, this model has limitations in handling data with outliers or poorly structured variables. The use of other techniques, such as density-based clustering algorithms or alternative modeling approaches, could serve as an alternative to improve the clustering quality and provide more optimal results.

Overall, this study makes a significant contribution to the effort to optimize BPJS health facility distribution, especially in addressing the challenges of geographic disparities and accessibility. The clustering results can serve as a basis for policy planning for equitable healthcare facilities in Indonesia and provide valuable insights for decision-making in distributing health facilities more efficiently. Of course, the implementation of these findings will heavily depend on support from various stakeholders, including the

government, BPJS, and the community, to collectively achieve better healthcare service equity.

CONCLUSION

The conclusion of this study is that the K-Means Clustering algorithm can be effectively used to optimize the distribution of BPJS health facilities in Indonesia. Based on the analysis results, it was found that the distribution of health facilities is uneven across regions, with some provinces and major cities having more comprehensive health facilities compared to remote areas. The grouping of health facilities based on location and type revealed disparities in access that need to be addressed, particularly in areas with limited healthcare facilities. This study also shows that there are four main clusters of health facilities that can be used as a basis for further distribution planning. This provides insights into areas that require additional health facilities or a more even distribution of resources. However, this study also identified several challenges, such as limitations in the data used and the need for more accurate and complete data collection. Overall, this research provides a strong foundation for planning a more optimal distribution of BPJS health facilities and equitable access to healthcare services across Indonesia. Nonetheless, the implementation of these findings requires cooperation between the government, BPJS, and the community to ensure a more efficient and fair distribution of healthcare facilities and improve the quality of services in areas that need them.

REFERENCES

Albar, C. (2023). *Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering Di Rsau Dr. Efram Harsana Lanud Iswahjudi Magetan*. Stikes Bhakti Husada Mulia Madiun.

Auditiyah, C. (2024). Pengelompokkan Daerah Rawan Demam Berdarah (Dbd) Di Jawa Timur Menggunakan Metode K-Means. *Estimasi: Journal Of Statistics And Its Application*, 205–215.

Ernawan, M. Z. T. O. F., & Andrian, R. B. K. (2024). *Optimalisasi Klasifikasi Stunting Dengan Metode Clustering Dan Rule Base Time Series (Rbt)*. Cv. Azka Pustaka.

Ferdinal, D., Nursukmi, I., Putra, R. R., & Wadisman, C. (2024). Prediksi Obat Kronis Penyakit Diabetes Melitus Menggunakan Metode Monte Carlo. *Jurnal Komputer Teknologi Informasi Dan Sistem Informasi (Juktisi)*, *3*(1), 665–672.

Hasugian, P. M., Kom, M., Hutahaean, H. D., Kom, M., Sinaga, B., Kom, M., Sriadhi, D.

R. S., Pd, M., Kom, M., & Silaban, D. R. S. (2022). *Teknologi Big Data Untuk Klasifikasi Status Desa Berdasarkan Indeks Desa Membangun (Idm) Melibatkan Algoritma K-Means Untuk Mendukung Program Kerja Kemendesa Pdtt*. Cattleya Darmaya Fortuna.

Ishmawati, N. U. R. (2024). *Analisis Clustering Umkm Sektor Perikanan Di Kabupaten Tangerang Menggunakan K-Modes Algorithm*. Fakultas Teknik Universitas Sultan Ageng Tirtayasa.

Komba, C., Bajuri, M. K., Beda, H., & Kaesmetan, Y. R. (2023). Optimisasi Stok Obat Bpjs Pada Bulan April 2023 Di Klinik Dompet Dhuafa Kupang Melalui Metode Topsis: Indonesia. *Jurnal Sosial Teknologi*, *3*(12), 1026–1034.

Mayola, L., Hafizh, M., & Syahputra, H. (2025). Klasterisasi Rumah Sakit Berdasarkan Kunjungan Pasien Menggunakan Algoritma K-Means: Data 2019-2023. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *7*(1), 15–21.

Syani, M., & Wahyudi, T. (2024). Klasterisasi Penggunaan Ban Dengan Cost Per Kilometer Terendah Pada Pt. Pl Menggunakan Metode K-Means. *Jurnal Indonesia: Manajemen Informatika Dan Komunikasi*, *5*(3), 2792–2800.