Research Horizon

ISSN: 2808-0696 (p), 2807-9531 (e)

Research Horizon
Volume: 06
Issue: 03
Year: 2026
Page: 1131-1144

Citation:
Putra, I. P., & Gunawan, H. (2026). Analysis of Indonesian migrant worker patterns using the K-Means clustering algorithm. *Research Horizon*, *6*(3), 1131-1144.

Article History:
Received: May 19, 2026
Revised: June 4, 2026
Accepted: June 22, 2026
Online since: June 25, 2026

Analysis of Indonesian Migrant Worker Patterns Using the K-Means Clustering Algorithm

Ihsan Pratama Putra[1*], Hendra Gunawan[1]

[1] Department of Informatics Engineering, STMIK IM, Bandung, Indonesia

[*] *Corresponding author: Ihsan Pratama Putra (icanpp6@gmail.com)*

Abstract

The increasing number of Indonesian migrant workers distributed across various destination countries has created a need for data-driven analysis to better understand migrant worker distribution patterns and support effective policy formulation. This study aims to analyze the characteristics and patterns of Indonesian migrant workers in Cianjur Regency using a data mining approach with the K-Means Clustering algorithm. The dataset used was obtained from Indonesian migrant workers placement data based on destination countries and processed through the Knowledge Discovery in Databases stages, including data selection, preprocessing, transformation, clustering, and evaluation. The clustering process was carried out using the K-Means algorithm, while the optimal number of clusters was determined using the Elbow Method and validated using the Silhouette Score. The results showed that the optimal number of clusters was 2 clusters with a Silhouette Score value of 0.849, indicating good clustering quality. The first cluster was dominated by destination countries with high numbers of migrant workers, while the second cluster consisted of countries with relatively lower numbers of migrant workers. These findings are expected to support data-driven decision-making for migrant worker placement policies.

Keywords

Clustering, Elbow Method, Indonesian Migrant Workers, K-Means, Silhouette Score.

1. Introduction

Indonesian Migrant Workers (Pekerja Migran Indonesia/PMI) play a significant role in Indonesia's economic development by contributing remittances and helping reduce domestic unemployment (Todaro & Smith, 2015). For many individuals in regions with limited employment opportunities, labor migration serves as a strategy to improve household income and living standards. Remittances provide substantial benefits for migrant workers' families, supporting household consumption, education, healthcare, and poverty alleviation (Dewandaru et al., 2019; Ukhtiyani & Indartono, 2020; Wahyuni & Sihaloho, 2022). These financial contributions also promote welfare improvements and socioeconomic changes in migrants' communities of origin (Mas'udah, 2020). Furthermore, economic motivations remain a primary driver of migration decisions, as workers seek higher wages and better employment prospects abroad than those available domestically (Anggara et al., 2024). Therefore, labor migration functions as both an economic and a social mobility strategy.

Cianjur Regency is one of the largest sources of Indonesian migrant workers in West Java. The high level of labor migration is driven by limited local employment opportunities, low household incomes, and expectations of better economic prospects abroad. Household economic conditions and social environments are important factors influencing individuals' decisions to work overseas (Hamidah, 2018). In addition, migration intentions are strengthened by social networks and information shared by former migrant workers, which provide insights into destination countries, job opportunities, and expected earnings (Spaan & Van Naerssen, 2020). Common destinations for migrant workers from Cianjur include Saudi Arabia, Taiwan, Hong Kong, Malaysia, and South Korea. Variations in labor demand across these countries contribute to different migration patterns, while employment opportunities, migration networks, and labor market demand significantly shape workers' mobility and destination choices (Angreni, 2022; Prianto et al., 2023).

Several variables are important in understanding the migration characteristics of Indonesian migrant workers, including destination country, number of workers, and placement year. The destination country reflects differences in labor demand, wage levels, and employment opportunities, while the number of workers indicates the concentration of migrant labor in specific countries. Placement year is also relevant because migration patterns may shift over time due to changes in economic conditions, labor regulations, and international cooperation. Labor migration is inherently dynamic and influenced by economic circumstances and job opportunities in destination countries (Anggara et al., 2024). Therefore, examining the relationships among these variables is essential for identifying migration trends and labor distribution patterns of Indonesian migrant workers from Cianjur Regency.

Studies on Indonesian migrant workers at the regional level still largely rely on descriptive approaches and primarily focus on social and economic impacts, such as remittances, welfare, and migration motivations. Limited attention has been given to computational methods that can uncover hidden patterns within migration data. Advances in data mining and machine learning offer opportunities for more systematic and objective analysis of migrant worker data (Frank & Hall, 2011). One widely used method is K-Means Clustering, which groups data based on similarities among observations. K-Means is considered simple, efficient, and effective for identifying homogeneous groups within datasets (Syakur et al., 2018). Consequently, clustering techniques have gained increasing relevance in socioeconomic research to support evidence-based policy formulation and decision-making.

The research gap in this study lies in the limited application of machine learning approaches in analyzing Indonesian migrant worker data, especially at the regional level. Previous studies have focused more on economic impacts, remittances, and welfare aspects without exploring migrant worker segmentation using clustering techniques (Praktikto et al., 2020; Mas'udah, 2020). In addition, studies specifically discussing labor migration patterns in Cianjur Regency are still limited despite the region being one of the largest PMI contributors in West Java. Therefore, this study offers novelty by implementing the K-Means Clustering algorithm to classify Indonesian migrant workers in Cianjur Regency based on destination country, number of employees, and placement year. Through this approach, migration patterns and labor distribution trends can be identified more objectively and systematically.

Thus, this study aims to classify Indonesian migrant workers in Cianjur Regency using the K-Means Clustering algorithm based on destination country, number of employees, and placement year variables. The results of this study are expected to provide useful insights for local governments and related institutions in developing data-driven migration policies, improving migrant worker protection programs, and designing employment strategies that are more adaptive to international labor market trends.

2. Literature Review

2.1. Indonesian Migrant Workers

PMI refers to Indonesian citizens who work abroad under specific employment contracts and receive legal protection from the government, as regulated in Law Number 18 of 2017 concerning the Protection of Indonesian Migrant Workers. PMI plays an important role in supporting the national economy through remittances, reducing unemployment, and improving household welfare in migrant-origin areas. In many regions, labor migration has become an alternative strategy to overcome limited employment opportunities and low income levels. However, migrant workers are also vulnerable to exploitation, discrimination, and legal issues in destination countries. Therefore, strengthening the protection system for Indonesian migrant workers is essential to ensure their rights, safety, and welfare throughout the migration process (Pattinussa & Rasyid, 2024; Anita, 2025).

The characteristics of migrant workers can be analyzed from demographic and occupational aspects such as age, gender, education level, marital status, work experience, and employment sector. These characteristics influence the type of work obtained, wage levels, and the destination countries chosen by migrant workers. Workers with lower educational backgrounds are generally employed in informal sectors such as domestic work and caregiving, while workers with higher educational qualifications tend to access formal employment sectors with better wages and protection. In addition, migration patterns are strongly influenced by economic and social factors. According to Agdifianti and Sundaya (2024), economic conditions, social environments, and demographic characteristics significantly affect the selection of destination countries for Indonesian migrant workers. Countries offering higher wages, stable employment opportunities, and strong labor demand tend to become the primary destinations for PMI. Wage disparities between domestic and foreign labor markets are among the dominant factors encouraging international labor migration among Indonesian workers (Sitoresmi & Suman, 2025; Yuniarti et al., 2025).

2.2. Data Mining

Data mining is the process of extracting useful information from large datasets to identify hidden patterns using statistical techniques, artificial intelligence, and machine learning methods. Han et al. (2012) explained that data mining is part of the Knowledge Discovery in Databases (KDD) process, which includes data selection, cleaning, transformation, and interpretation to produce meaningful information for decision-making. The development of big data and machine learning technologies has increased the use of data mining in various fields, including economics, business, healthcare, and social analysis (Maimon & Rokach, 2014). Through data mining techniques, complex datasets can be processed more efficiently to identify trends, relationships, and patterns that are difficult to detect manually (Larose, 2005).

In this study, data mining is used to analyze the characteristics and patterns of Indonesian migrant workers in Cianjur Regency based on variables such as age, gender, education level, destination country, and employment sector. This approach enables the identification of classification, grouping, and prediction patterns systematically and objectively. One commonly used technique is clustering, which groups data based on similarities in characteristics among data objects. According to Oyewole and Thopil (2023), clustering techniques are effective in identifying hidden structures and segmentation patterns within large datasets, making them valuable for data-driven analysis and policymaking processes. Furthermore, clustering methods have increasingly been utilized in labor market and migration studies to identify regional characteristics and employment distribution patterns more effectively (Gružauskas et al., 2021).

2.3. K-Means Clustering

K-Means is one of the unsupervised learning algorithms used to group data into several clusters based on similarities in data characteristics. According to MacQueen (1967), the K-Means method aims to minimize the distance between data points within the same cluster so that objects in one group have more similar characteristics compared to those in other groups. K-Means is widely used in data mining because it is simple, efficient, and capable of processing large datasets effectively. The algorithm works by determining the number of clusters (K), selecting initial centroids, calculating distances between data points and centroids, and repeatedly updating cluster memberships until stable cluster formations are obtained.

In this study, the K-Means algorithm was applied to classify Indonesian migrant worker data in Cianjur Regency based on characteristics such as age, education level, type of work, and destination country. This method was chosen because it can identify hidden patterns and group migrant workers with similar characteristics systematically and objectively. The clustering process used the Euclidean Distance method to measure the distance between data points and cluster centroids. Euclidean Distance is commonly used in K-Means because it calculates the shortest distance between two points in multidimensional space, enabling more accurate grouping results. Through this approach, labor migration patterns and PMI segmentation can be analyzed more effectively to support data-driven policymaking. In addition, K-Means clustering has been widely applied in socioeconomic and demographic studies because it is capable of identifying distribution patterns and hidden structures within complex datasets (Sasmita et al., 2023; Fujiansyah, 2025; Unukić & Nater Drvenkar, 2025).

$$(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Description:
$d\ (x, y)$ = The distance between the data and x *cemtroid* y
x_i = The value of the data attribute to $- i$
y_i = Value *Centroid* on the attribute to $- i$
n = number of attributes

SSE Formula:

$$SSE = \sum_{\mathrm{i}=1}^{\mathrm{k}} \sum_{x \in C_i} (x - C_i)^2$$

3. Methods

This study employed a quantitative research approach with the application of data mining techniques to analyze the characteristics and patterns of Indonesian Migrant Workers in Cianjur Regency (Hair et al., 2019). The research aimed to identify migrant worker grouping patterns based on destination countries and the number of migrant workers using the K-Means Clustering algorithm. The dataset used in this study was obtained from official PMI placement data containing information related to destination countries, the number of migrant workers, and the year of placement. These variables were selected because they represent important aspects in describing labor migration patterns and the distribution of Indonesian migrant workers abroad.

The variables analyzed in this study consisted of destination country as a categorical variable representing the placement country of PMI, number of employees as a numerical variable representing the total number of migrant workers in each destination country, and year as a numerical variable indicating the placement year of PMI. The collected data were processed using the Knowledge Discovery in Databases (KDD) framework, which is a systematic approach for extracting patterns and meaningful information from large datasets. The KDD process in this study consisted of several stages, namely data selection, data cleaning, transformation, clustering, and evaluation.

The first stage was data selection, which involved selecting relevant variables from the PMI database, including destination country, number of migrant workers, and placement year, to analyze labor migration patterns in Cianjur Regency. The next stage was data cleaning, which aimed to improve data quality by handling missing values, duplicate records, and inconsistencies to ensure accurate and reliable clustering results. After cleaning, the data transformation stage was conducted using the StandardScaler method to standardize numerical variables. This process was necessary because the K-Means algorithm uses Euclidean Distance calculations that are sensitive to differences in data scales. Standardization transformed the data into a uniform scale, enabling a more optimal and balanced clustering process.

After the transformation process, clustering was conducted using the K-Means algorithm. This method groups data based on similarities in characteristics by calculating the distance between data points using the Euclidean Distance approach. The number of clusters (K) was determined to classify PMI data into several groups with similar migration patterns and characteristics (Everitt et al., 2011). The clustering results were expected to reveal patterns related to destination countries and labor distribution trends among Indonesian migrant workers in Cianjur Regency.

The final stage was evaluation, which aimed to measure the quality and effectiveness of the clustering results. In this study, the Elbow Method was used to determine the optimal number of clusters, while the Silhouette Score was applied to evaluate the level of similarity within clusters and the separation between clusters. A higher Silhouette Score indicates better clustering quality because it reflects high homogeneity within clusters and strong heterogeneity between different clusters. Through these evaluation techniques, the clustering results can be validated and interpreted more effectively to support data-driven analysis and policymaking related to Indonesian migrant workers.

4. Results

The initial stage of the research is carried out through a data selection process to determine attributes that are relevant to the research objectives. The variables used in the clustering process include the destination country of placement of migrant workers, the number of Indonesian migrant workers, and the year of placement. The selection of attributes was carried out so that the analysis process was more focused on identifying PMI distribution patterns based on the work destination country. Next, the data preprocessing stage is carried out to improve the quality of the dataset before the clustering process is carried out. The preprocessing stage includes the removal of missing values, duplicate data, and data that has a value of the number of workers less than or equal to zero. This stage is important because the quality of the data greatly affects the results of the grouping in the K-Means algorithm.

After the data cleansing process is complete, the data is then grouped by destination country to obtain the total number of PMIs in each country. The results of the data aggregation show that the distribution of the number of PMIs between destination countries has a significant difference. Some countries have very high PMIs, while others have relatively low PMIs. A good preprocessing stage helps to improve the stability of the clustering model so that the grouping results become more accurate and representative of the actual conditions.

The data transformation stage is carried out using the standardization method with StandardScaler. This transformation aims to equalize the scale of numerical data so that the distance calculation process in the K-Means algorithm is not affected by differences in the range of values between attributes. The K-Means algorithm uses Euclidean Distance calculation to determine the proximity between data and the centroid cluster. Therefore, attributes with a large range of values can dominate the clustering process if normalization or standardization is not carried out first. Through the standardization process, the data on the number of PMIs is changed to a distribution with an average of close to zero and a standard deviation of one. Thus, each data point has a balanced contribution in the process of forming clusters. This transformation stage also helps improve the performance of the clustering model and results in more optimal cluster separation.

The determination of the optimal number of clusters is carried out using the Elbow method by calculating the Within Cluster Sum of Squares (WCSS) value on several variations in the number of clusters. WCSS is used to measure the level of homogeneity of data in a cluster. The smaller the WCSS value, the closer the distance between the data in the same cluster. In this study, the number of clusters from $k = 2$ to $k = 6$ was tested. Based on the test results, the Elbow graph shows a significant decrease in the value of WCSS up to a certain point and begins to slope after the 2nd cluster count.

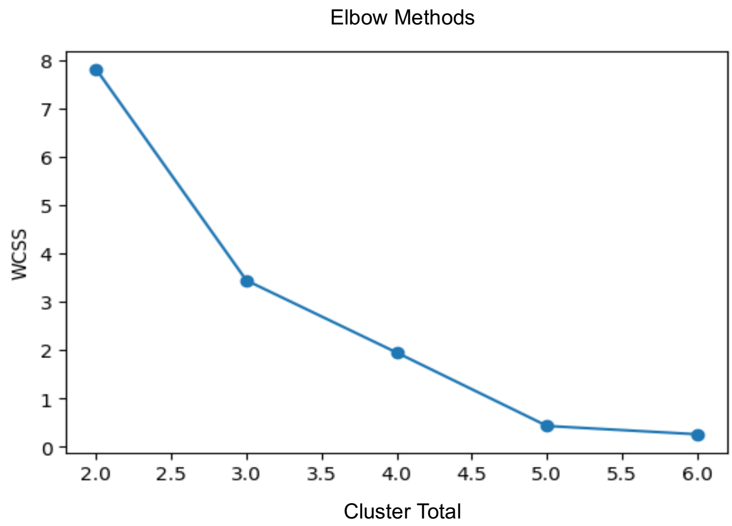

Figure 1. Elbow Method Chart

Based on Figure 1, the elbow point is seen at the value k = 2, so that the optimal number of clusters selected in this study is as many as 2 clusters. The selection of the number of clusters was carried out because, after that point, the decrease in the WCSS value was not too significant. These results show that the division of data into two groups is quite capable of representing the distribution pattern of PMI based on the country of work destination.

After the optimal number of clusters is obtained, the next stage is to evaluate the quality of clustering using the Silhouette Score method. This method is used to measure the degree of similarity of data within the same cluster as well as the degree of difference between clusters. The Silhouette Score value is in the range of -1 to 1. The closer the value is to 1, the better the quality of clustering because it shows that the data in one cluster has a high level of similarity and has a considerable distance from other clusters.

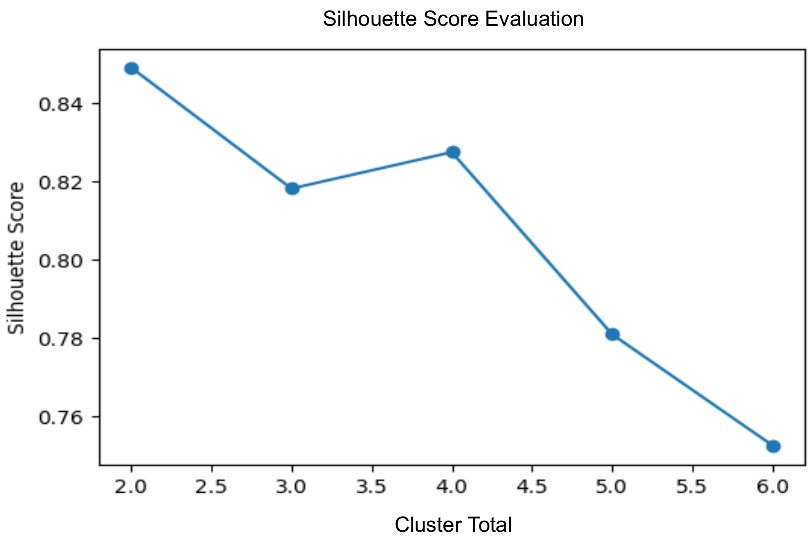

Figure 2. Silhouette Score Graph

Based on Figure 1 results, a Silhouette Score of 0.849 was obtained. A Silhouette Score of 0.849 indicates that the clustering results are of excellent quality. This indicates that the process of grouping PMI data using the K-Means algorithm is able to produce a fairly clear cluster separation and minimal overlap between groups. In addition, the high evaluation value shows that the data in each cluster has relatively homogeneous characteristics, so that the clustering results can be used as a basis for more objective PMI pattern analysis.

The clustering process was carried out using the K-Means algorithm with an optimal number of clusters of 2 clusters based on the results of the Elbow Method and Silhouette Score evaluations. Grouping is carried out on PMI data based on the number of migrant workers in each destination country. The clustering results show that PMI data is divided into two main groups with different characteristics. The first cluster is a group of destination countries with a relatively low number of PMIs, while the second cluster consists of destination countries with a high number of PMIs.

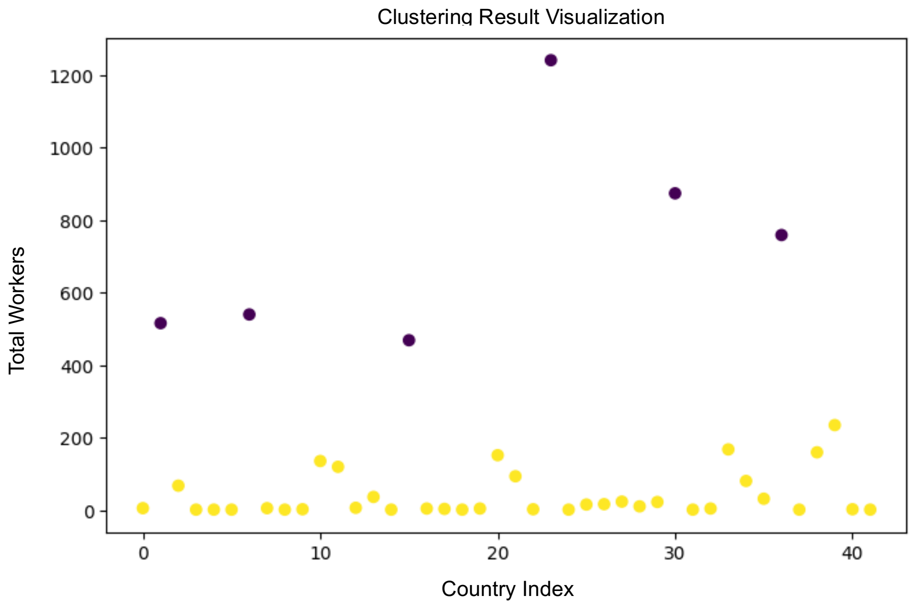

Figure 3. Visualization of K-Means Clustering Results

The clustering visualization shows in Figure 3 that there is a fairly clear separation of data between one cluster and another. Destination countries with a high number of migrant workers tend to form separate groups because they have different data distribution characteristics compared to countries with a low number of migrant workers. The cluster with a high number of migrant workers shows that the country has a great attraction for Indonesian migrant workers. Factors that affect the high number of migrant workers in certain countries can be caused by the high demand for labor, better wage levels, easy access to job placement, and the existence of a social network of migrant workers that has been formed beforehand.

Meanwhile, clusters with low numbers of migrant workers indicate that the destination country has a relatively lower placement rate of migrant workers. This condition can be influenced by labor regulation factors, limited job opportunities, and low public interest in certain destination countries. The results of this clustering provide an idea that the distribution pattern of PMI is not evenly distributed across all destination countries. Therefore, the results of this research can be used by local governments and related agencies in developing a more targeted placement and protection strategy.

Table 1. Characteristics of Clustering Results

Cluster	Characteristics
Cluster 0	Destination countries with relatively low PMI numbers
Cluster 1	Destination countries with a high number of PMIs

Based on Table 1 of the grouping, the government can focus more on supervising, protecting, and coaching migrant workers in destination countries with a high number of placements. In addition, the clustering results can also be used as a basis for evaluating the PMI placement policy to be more in line with the conditions and needs of Indonesian migrant workers.

5. Discussion

Based on the results of the research, the K-Means Clustering algorithm was successfully applied to analyze the distribution patterns of Indonesian migrant workers in Cianjur Regency based on destination country, number of employees, and placement year. This study implemented the Knowledge Discovery in Databases (KDD) approach, which includes data selection, preprocessing, transformation, clustering, and evaluation stages. According to Han et al. (2012), the KDD process is effective for extracting meaningful information and hidden patterns from large datasets systematically. Through this approach, PMI data could be processed more objectively to identify labor migration patterns and country placement trends. The use of data mining techniques in this study also demonstrates the growing importance of computational methods in socioeconomic analysis and policymaking (Tan et al., 2016).

The clustering process using the K-Means algorithm produced an optimal number of two clusters based on the Elbow Method analysis. This finding is consistent with Syakur et al. (2018), who explained that the integration of the Elbow Method with K-Means clustering is effective in determining the optimal number of clusters and improving clustering quality. The evaluation results showed a Silhouette Score value of 0.849, indicating excellent clustering performance with high intra-cluster similarity and clear inter-cluster separation. According to Jain (2010), a high Silhouette Score reflects that the clustering model is capable of representing data characteristics effectively and minimizing overlap between clusters. This result indicates that the PMI data in Cianjur Regency have relatively distinct grouping patterns based on destination country and labor distribution.

The clustering results identified two main groups of destination countries. The first cluster consisted of countries with high numbers of Indonesian migrant workers, while the second cluster represented countries with relatively lower PMI placements. Countries included in the high-cluster category generally offer better wage levels, larger employment opportunities, and stronger labor demand compared to other countries. This finding supports the study of Fadhilah and Sundaya (2023), which stated that economic opportunities and wage differences are major factors influencing the selection of destination countries for Indonesian migrant workers. Similarly, Anggara et al. (2024) emphasized that higher income expectations and limited domestic job opportunities are dominant motivations for Indonesian workers seeking employment abroad.

The concentration of PMI placements in certain countries also reflects the importance of remittances for household welfare and regional economic development. Mas'udah (2020) explained that remittances from migrant workers significantly influence lifestyle changes and socioeconomic conditions in migrant-origin communities. In addition, Pratikto et al. (2020) stated that remittances supported by strong social capital can improve household economic resilience. This is also supported by Ukhtiyani and Indartono (2020), who found that migrant worker remittances contribute positively to Indonesia's economic growth. In the context of Cianjur Regency, the high number of PMIs indicates that international labor migration has become an important economic strategy for households facing limited local employment opportunities.

Furthermore, the findings of this study have practical implications for government institutions and policymakers. The clustering results can help related agencies identify priority destination countries, labor distribution patterns, and migrant worker concentrations more effectively. This information can support the formulation of more targeted policies regarding PMI placement, labor training, supervision, and protection. Pattinussa and Rasyid (2024) emphasized that strengthening migrant worker protection systems is essential to ensure workers' rights and welfare during the migration process. Therefore, the implementation of data mining and clustering approaches in PMI analysis can become an important tool for supporting evidence-based policymaking and improving migrant worker governance in Indonesia.

6. Conclusion

This study successfully applied the K-Means Clustering algorithm to analyze the distribution patterns of Indonesian Migrant Workers in Cianjur Regency based on destination country, number of employees, and placement year. Using the Knowledge Discovery in Databases (KDD) approach, the research demonstrated that data mining techniques can effectively identify labor migration patterns systematically and objectively. The evaluation results showed that the optimal clustering model consisted of two clusters, supported by a high Silhouette Score value, indicating good clustering quality with clear separation between groups. The findings revealed that several destination countries have significantly higher concentrations of migrant workers due to factors such as employment opportunities, wage levels, and labor demand.

The results of this study provide practical implications for local governments and related institutions in developing data-driven policies related to migrant worker placement, supervision, and protection. By understanding PMI distribution patterns, policymakers can design more targeted training programs, improve labor protection systems, and strengthen migration governance, particularly for destination countries with high placement rates. In addition, the study highlights the importance of integrating machine learning and data mining approaches into socioeconomic analysis to support evidence-based decision-making.

However, this study has several limitations. The analysis only used a limited number of variables, namely destination country, number of employees, and placement year, which may not fully represent the complexity of labor migration patterns. Furthermore, the study only applied the K-Means algorithm, which may produce different results compared to other clustering methods. Therefore, future research is recommended to include additional variables such as age, gender, education level, and employment sector to provide more comprehensive analysis results. Further studies may also compare the performance of other clustering methods, such as DBSCAN, Fuzzy C-Means, and Hierarchical Clustering, to improve clustering accuracy and generate deeper insights into Indonesian labor migration patterns.

References

Agdifianti, S., & Sundaya, Y. (2024). Analisis ekonomi pekerja migran Indonesia dalam memilih jenis pekerjaan pada BP3MI Provinsi Jawa Barat. *Dinamika Ekonomi*, *7*(4), 31-38. https://doi.org/10.29313/JDE.V15I1.3082.

Anggara, R., Mulyana, S., Gayatri, G., & Hafiar, H. (2024). Understanding the motivations of being Indonesian migrant workers. *Cogent Social Sciences*, *10*(1), 233-244. https://doi.org/10.1080/23311886.2024.2333968.

Angreni, D. K. D. (2022). Social network analysis of Indonesia migrant workers in Hong Kong: Study social integration between health care and KOTKIHO. *International Journal of Public Administration Studies*, *2*(1), 07-13. https://doi.org/10.29103/ijpas.v2i1.7854.

Anita, Y. (2025). Empowerment of migrant workers to prevent illegal employment: A review of government policies. *Research Horizon*, *5*(2), 373-384. https://doi.org/10.54518/rh.5.2.2025.540.

Dewandaru, B., Rahmadi, A. N., & Syaâ, E. H. (2019). Pemanfaatan remitansi pekerja migran Indonesia serta peran usaha pekerja migran Indonesia purna untuk pembangunan desa asal. *Warmadewa Economic Development Journal (WEDJ)*, *2*(2), 44-50. https://doi.org/10.22225/wedj.2.2.1297.44-50.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). London: Wiley.

Fadhilah, A. N., & Sundaya, Y. (2023). Analisis ekonomi pekerja migran Indonesia dalam memilih negara tujuan pada BP3MI Jabar. *Jurnal Riset Ilmu Ekonomi dan Bisnis*, *6*(4), 111-116. https://doi.org/10.29313/jrieb.v3i2.2856.

Frank, E., & Hall, M. A. (2011). *Data mining: practical machine learning tools and techniques.* New York: Morgan Kaufmann.

Fujiansyah, D. (2025). Segmentation of inclusive economic growth profiles across provinces in Indonesia using a clustering approach. *Indonesian Journal of Social Economics and Agricultural Policy*, *1*(1), 33-49. https://doi.org/10.70895/ijseap.v1i1.64.

Gružauskas, V., Čalnerytė, D., Fyleris, T., & Kriščiūnas, A. (2021). Application of multivariate time series cluster analysis to regional socioeconomic indicators of municipalities. *Real Estate Management and Valuation*, *29*(3), 39-51. https://doi.org/10.2478/remav-2021-0020.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis.* New York: Pearson.

Hamidah, C. N. (2018). Sirkulasi keputusan dan dampak menjadi pekerja migran: studi etnografi proses pengambilan keputusan menjadi pekerja migran Indonesia. *Jurnal Ketenagakerjaan*, *13*(1), 559-666.

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques.* Waltham: Morgan Kaufmann Publishers.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011.

Larose, D. T. (2005). An introduction to data mining. In *Traduction et adaptation de Thierry Vallaud.* London: John Wiley.

MacQueen, J. (1967). Multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281-297). Oakland, CA: University of California Press.

Maimon, O. Z., & Rokach, L. (2014). *Data mining with decision trees: theory and applications* (Vol. 81). Cham: World Scientific.

Mas'udah, S. (2020). Remittances and lifestyle changes among Indonesian overseas migrant workers' families in their hometowns. *Journal of International Migration and Integration*, *21*(2), 649–665. https://doi.org/10.1007/s12134-019-00676-x.

Oyewole, G. J., & Thopil, G. A. (2023). Data clustering: application and trends. *Artificial Intelligence Review*, *56*(7), 6439-6475. https://doi.org/10.1007/s10462-022-10325-y.

Pattinussa, J. M. Y., & Rasyid, A. G. (2024). Enhancing protection of Indonesian migrant workers: A case study of Erwiana. *Jurnal Transformasi Global*, *11*(1), 62–74. https://doi.org/10.21776/ub.jtg.011.01.5.

Pratikto, R., Yazid, S., & Dewi, E. (2020). Enhancing the role of remittances through social capital: Evidence from Indonesian household data. *Asian and Pacific Migration Journal*, *29*(1), 30–54. https://doi.org/10.1177/0117196820920401.

Prianto, A. L., Amri, A. R., & Ajis, M. N. E. (2023). Governance and protection of Indonesian migrant workers in Malaysia: a study on policy and innovation network. *JSEAHR*, *7*(4), 214-222.

Sasmita, N. R., Khairul, M., Sofyan, H., Kruba, R., Mardalena, S., Dahlawy, A., Apriliansyah, F., Muliadi, M., Saputra, D. C. E., Noviandy, T. R., & Maula, A. W. (2023). A statistical clustering approach: Mapping population indicators through probabilistic analysis in Aceh Province, Indonesia. *Infolitika Journal of Data Science*, *1*(2). https://doi.org/10.60084/ijds.v1i2.130

Sitoresmi, L. N., & Suman, A. (2025). Determinants of labor international migration in Java. *Journal of Development Economic and Social Studies*, *4*(4), 1365-1379. https://doi.org/10.21776/jdess.2025.04.4.21.

Spaan, E., & Van Naerssen, T. (2020). Migration decision-making and migration industry in the Indonesia–Malaysia corridor. In *Exploring the Migration Industries* (pp. 138-153). London: Routledge.

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP Conference Series: Materials Science and Engineering* (Vol. 336, No. 1, p. 012017). London: IOP Publishing. https://doi.org/10.1088/1757-899X/336/1/012017.

Tan, P. N., Steinbach, M., & Kumar, V. (2016). *Introduction to data mining*. Chennai: Pearson Education India.

Todaro, M. P., & Smith, S. C. (2015). *Economic development (12th edition)*. Essex: Essex Pearson Education Limited.

Ukhtiyani, K., & Indartono, S. (2020). Impacts of Indonesian economic growth: Remittances migrant workers and FDI. *JEJAK: Jurnal Ekonomi dan Kebijakan*, *13*(2), 280–291. https://doi.org/10.15294/jejak.v13i2.23543.

Unukić, I., & Nater Drvenkar, N. (2025). Clustering regional competitiveness in Central and Eastern Europe: Insights from the k-means method. *Ekonomski vjesnik: Review of Contemporary Entrepreneurship, Business, and Economic Issues*, *38*(2), 373-386. https://doi.org/10.51680/ev.38.2.12.

Wahyuni, & Sihaloho, M. (2022). Hubungan remitan ekonomi dengan tingkat kesejahteraan rumah tangga pekerja migran Indonesia:(Kasus: Desa Galak, Kabupaten Ponorogo, Jawa Timur). *Jurnal Sains Komunikasi dan Pengembangan Masyarakat*, *6*(2), 202-218. https://doi.org/10.29244/jskpm.v6i2.703.

Yuniarti, D., Astuti, H., Triatmojo, F. A., Nasir, M. S., & Lunku, H. S. (2025). Quantifying the poverty paradox: Indonesian labor migration. *Jurnal Ekonomi Pembangunan: Kajian Masalah Ekonomi dan Pembangunan*, *26*(2), 233-249. https://doi.org/10.23917/jep.v26i2.13174.

Acknowledgment
We gratefully acknowledge the contributions of individuals who supported the completion of this article.

Funding Information
This research did not receive any funding.

Conflict of Interest Statement
The authors declare that there is no conflict of interest.

Ethical Approval and Originality Statement
Ethical approval was obtained for this study. The manuscript represents original work and has not been previously published, nor is it under consideration by another journal.

Data Disclosure Statement
The data that support the findings of this study are available from the corresponding author upon reasonable request.