

Clustering Village Development in West Java Province on the Condition of Developing Village Strata Using K-Means Algorithm

Nurahman^{1*}, Nurani Aisyah Tanjung¹

¹ Information Systems, Faculty of Computer Science, Darwan Ali University, Sampit, Indonesia.

Received: October 17, 2023

Revised: November 29, 2023

Accepted: September 25, 2023

Published: September 30, 2023

Corresponding Author:

Nurani Aisyah Tanjung

Nuraniaisyah477@gmail.com

DOI: [10.29303/jppipa.v9iSpecialIssue.5937](https://doi.org/10.29303/jppipa.v9iSpecialIssue.5937)

© 2023 The Authors. This open access article is distributed under a (CC-BY License)



Abstract: Villages are important units in the socio-economic structure of a country, and an in-depth understanding of the factors that influence village growth and welfare is essential. To facilitate the local government in handling the equitable distribution of village needs, it is necessary to cluster or group villages. The purpose of this research is to assist the government in clustering certain villages into several clusters, making it easier to monitor and procure village needs within the West Java Provincial government. Clustering is done using the K-Means algorithm. The application of the K-Means Algorithm by determining the Cluster value is 96. The results showed that each cluster has its own membership number. Cluster 0 consists of 8 villages, Cluster 1 consists of 12 villages, and Cluster 2 consists of 8 villages. Furthermore, in testing the performance of the K-Means algorithm by dividing into 96 clusters, the Davies Bouldin Index value is -0.996. From the results of data processing and analysis, it is necessary to provide assistance and procurement of village needs for villages with low clusters by the West Java provincial government.

Keywords: Clustering; Data mining; K-Means algorithm

Introduction

Villages are an important part of the Indonesian economy. The Government of Indonesia is working to improve the quality of life of rural communities, including access to education and health services. Although there are still challenges ahead, village development gives hope for Indonesia to live better and develop sustainably.

West Java is an Indonesian province in the western part of Java Island. All villages have different social levels. In 2023, there were 1828 villages with independent strata, 2553 villages with developed strata, and 930 villages still in the developing village category. (DPM-Desa Provinsi Jawa Barat, 2023) In this study, the data source used to determine the development of villages in West Java must manually collect data on various important factors, such as environmental health

infection, energy consumption, healthy family index, and village development index.

In this process, collaboration with government and communities is necessary to collect and study data thoroughly. The challenges faced in collecting data to manage the necessary information pose difficulties when using legacy systems. This process requires a lot of time and effort. However, limited data sources, data privacy issues, changes in data availability, and various data formats are now no longer a big challenge to do so. Thanks to the advancements in computer science which are very important.

Computer science has developed into one of the most useful fields of science and technology along with the increasing demand for data to manage information. (Kouhi Esfahani et al., 2019) Various advances in computer science include data mining. Data mining is a branch of computer science that uses algorithms and data analysis techniques to cluster raw data for better

How to Cite:

Nurahman, N., & Tanjung, N.A. (2023). Clustering Village Development in West Java Province on the Condition of Developing Village Strata Using K-Means Algorithm. *Jurnal Penelitian Pendidikan IPA*, 9(SpecialIssue), 299-306. <https://doi.org/10.29303/jppipa.v9iSpecialIssue.5937>

decision making. clusters are one of the most frequently used processes in unsupervised learning (Economou, 2023), Clustering is the task of grouping observations into clusters (Chen et al., 2020), K-Means Clustering Algorithm is an effective solution (Okta Jaya Harmaja et al., 2023), The advancement of information technology is now growing rapidly and is almost covers all areas of life (Alhapizi et al., 2020), required a technique so that the processing results or information obtained obtained is appropriate (Virgo et al., 2020), Data mining involves the process of extracting data from various sources (Sioyong, 2023).

One of the interesting data mining techniques is clustering. Agarwal & Mehta, (2016) is one of the methods in data mining. A partitioning method that divides all N objects into K clusters in such a way as to provide high intra-cluster similarity and low inter-cluster similarity. Vats & Sagar, (2019) and one of the most simple and popular algorithms for similar data aggregation techniques is K-Means.

In previous research, the K-means algorithm has also been carried out in research on the removal of irrelevant data based on the classification of human activities in smart home sensor networks (Pattamaset & Choi, 2020). The results showed that the IDEK approach, which includes a cluster-based algorithm embedded in the cluster head, can improve sensor network data aggregation by removing irrelevant data, because the cluster head acts as a local processor that preprocesses data and removes irrelevant data.

In research on iterative algorithms for optimal variable weighting (Zhang et al., 2019). It is known from the research results that the weighted Kmeans algorithm can improve performance in non-homogeneous and non-spherical cases by suppressing noise variables and transferring non-spherical space into spherical space with appropriate variable weights.

This research has a connection to previous research which is used as reference material. By confirming and strengthening the results of previous studies, this research makes an additional contribution to the literature. By comparing the existing data with previous research, it is possible to identify similar trends and patterns in both studies. This research also has some gaps to previous research as shown in table 1 which shows the position of related research in this article.

The literature review taken in this study is on the topic of Urban flood risk assessment based on DBSCAN and K-means clustering algorithm (Li et al., 2023) which in previous studies conducted pollination clustering (FPA-C), bat clustering (BA-C) and firefly clustering (FFA-C). In the research An Improved K-Means Clustering for Segmentation of Pancreatic Tumor from CT Images (Reena Roy & Anandha Mala, 2023) This research performs dynamic particle swarm optimization

(DPSO) clustering and K-means clustering, in the research Nonparametric K-means algorithm with applications in economic and functional data (Feng & Zhang, 2022) performs raw pixel clustering using different image scaling percentages, in the research Newsgroup topic extraction using term-cluster weighting and Pillar K-Means clustering (Adinugroho et al., 2022) performs clustering of social media, online news portals, and newsletters, while in the research BIM performance assessment system using a K-Means clustering algorithm (Kim et al., 2021) performs clustering Similar days of the year, sunny days, cloudy days, and rainy days.

By using the K-Means algorithm applied to recognise a similarity of data (Zhang et al., 2019) This research is also to cluster similar data into various groups that have comparable characteristics, which allows for more in-depth analysis and a better understanding of the patterns that may exist in the data. Clustering uses K-means as a clustering technique, because the K-means algorithm can handle it effectively for large data sets (Varanasi & Tripathi, 2019). The k-means algorithm is based on partitioning and mainly finds similarities by calculating distances. In K-means technique, the data is randomly partitioned and k centre points are selected. Refine the partition according to the distance between the k centre points and the remaining k data. Calculating the Euclidean distance, the clustering technique assigns each data to the nearest centre point calculated based on Eq (Kamalha et al., 2018).

Through this scientific publication, it is the obligation of researchers or scientists to announce the results, findings, conclusions, and implications of the results of research or review to the public at large (Gunawan et al., 2019). To assist the government in clustering villages in West Java into several clusters that make it easier for related parties to carry out development and equitable development of villages towards independent village strata in the province of West Java, Indonesia.

Method

The research methodology is a framework that contains the stages used in research. A research framework is the basic structure or frame of reference (Yolanda, 2023). The research stage carried out in this study is to find problems, analyse problems, study literature, and collect data, then apply the Knowledge Discovery in Databases (KDD) approach. The stages of the research conducted are depicted in Figure 1.

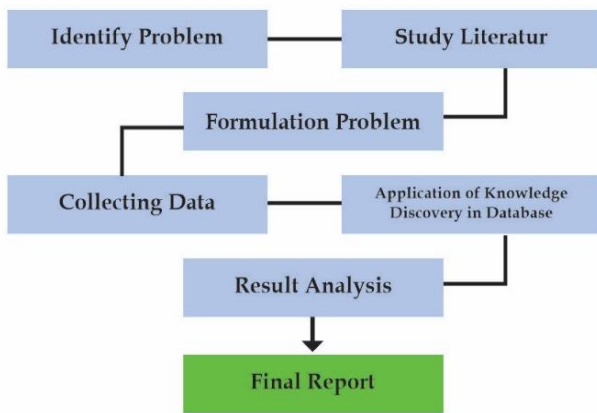


Figure 1: Research Stages

The following is a description of the research stages:

1. The purpose of problem identification is to provide some questions to researchers so that researchers can find ways to solve these problems in the future, in other words, good problem identification will describe the problems that exist in the research topic.
2. In this study, researchers conducted searches by reviewing books and literature related to current issues.
3. After identifying the problem and studying the literature, the next step is problem formulation. This problem is carried out to obtain a problem formulation that will be the focus of this research.
4. The data collection process was carried out by observing the official website of the West Java Provincial Government of West Java Province. Then directed to the West Java Village Portal 2023 page.
5. Once the data was collected, the Knowledge Discovery in Database (KDD) methodology was utilised. There are five important steps involved in data processing in this study; namely selection, preprocessing, transformation, data mining, and interpretation evaluation. It can be seen in figure 2.

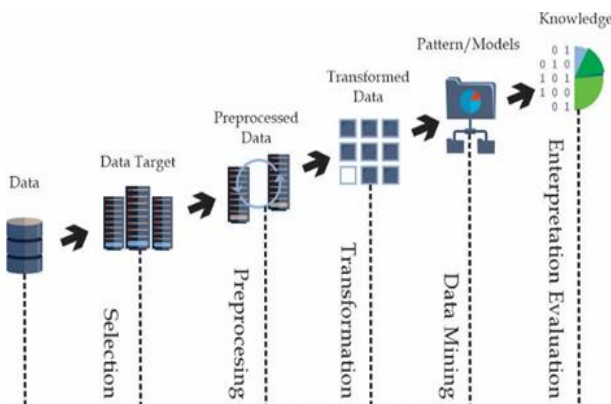


Figure 2 : Knowledge Discovery in Database (KDD) Methodology

6. After processing data or applying the Knowledge Discovery Database (KDD) approach, the next step is to analyse the results of data processing. Analyse by understanding the results of data processing by reviewing some previous research.
7. The preparation of the final report is done by writing a scientific article based on observation, data processing and analysis of results.

Knowledge Discovery In Database (KDD) in this study to find out knowledge in decision making. Knowledge Discovery in Database (KDD) is a model for obtaining knowledge from existing databases (Filki, 2022), that exists. The stages in the Knowledge Discovery in Database (KDD) methodology in this study are shown in Figure 2. The first stage is Selection, Preprocessing, Transformations, Data Mining and finally Interpretation Evaluation.

In the K-Means algorithm, the dataset analysed in this study does not have a label so it is appropriate in this study to cluster using the K-Means algorithm. In the K-Means algorithm, each observation belongs to the group that has the closest mean and the group tries to find its centre. (Liu & Chen, 2019). The number of groups in the data is equal to the number of refinement iterations performed by the algorithm. This research performs the mining process using the K-Means algorithm by applying equation (1).

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

The clustering process begins with identifying the data to be clustered (Fuss et al., 2016) After the data is identified, the equation can be applied to calculate the distance value between data x and data y . where to find out x_i is the i -th test data and y_i is the i -th training data. Then proceed to find the centroid point value of the cluster to K . To find the centroid point value of a cluster to K can be done using the following equation (2)

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i \tag{2}$$

μ_k is the centroid point value of cluster K , then the symbol N_k is used to represent the value of the volume of data in k clusters, and then the symbol x_i is used to represent the value. The i -th data in the k th cluster. Figure 3 shows the stages of data processing using the K-Means algorithm. The data set has the final data processing will be done using the K-Means algorithm. The number of K values that will be formed will determine the number of clusters, then find the centroid point value to group the data based on the number of clusters determined. If the centroid point changes,

reselect the centroid point until you know that the centroid point has not changed.

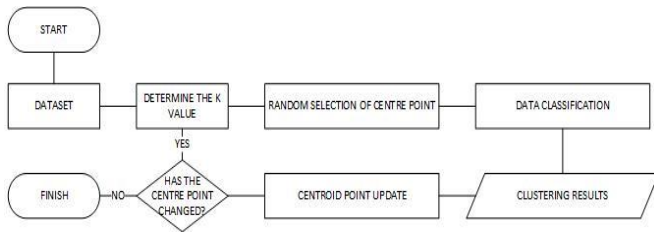


Figure 3. K-Means Algorithm Stages

Description of each stage of the K-Means algorithm:

1. Start by applying the K-Means algorithm calculation
2. The dataset used is data on developing villages in West Java province.
3. Determine the value of K to determine the group or cluster the school will form.
4. Randomly select a centre of mass point and count each data set into a cluster considering.
5. Grouping the data so that K clusters are formed based on the centroid point of each cluster.
6. Data clustering results based on iterations
7. Update the centroid point value and determine if the centroid point has changed. If there is no change The calculation has stopped

Testing the performance of the K-means algorithm is by finding the value of the Davies Boildin Index. The Davis Boildin index is located during the data mining process. The search for the Davies Boildin index value is clustered using the K-Means algorithm according to the predetermined K value. The clusters found are then tested using the Davies Boildin Index measurement principle. The smaller the Davies Boildin Index value, the more optimal the cluster scheme. The formula of Davies Boildin Index calculation can be presented in equation (3).

$$DBI = \frac{1}{k} \sum_{i=1}^k \{R_i\}$$

By

$$R_i = \frac{S_i + S_j}{d_{ij}}, \quad i = 1, \dots, k, i \neq j, R_{ij} = \frac{S_i + S_j}{d_{ij}}$$

And

$$S_i = \left[\frac{1}{n} \sum_{x \in n_i} d^2(x, v_i) \right] \frac{1}{2} \tag{3}$$

In the process of finding the Davies Boildin Index value contained in equation (3) that the symbol k shows the value of the number of clusters and R_{ij} is a measure of similarity between the values n_i and n_j . The S_i symbol shows the size of the i -th cluster dispersion, where $i = 1, 2, \dots, k$. The symbol d_{ij} indicates the distance between the centroid of the i -th cluster and the centroid of the j -

th cluster ($d_{ij} = d_{ji}$). The symbol n_i indicates the number of members of the i -th cluster, $i = 1, 2, \dots, k$. Then V_i indicates the cluster centroid value of n_i .

Result and Discussion

In the research methodology, the research was conducted in several stages. Firstly, the problem was identified by looking at the comparative strata of developing villages in West Java province. The observation shows the problem of differences in the healthy family index, energy consumption intensity, environmental health inspection, and village development index.

In addition, a literature study was conducted to support the scientific and theoretical findings. After the case was known in its entirety, the problem was formulated to improve the focus of the research. For example, knowing which villages can enter the developed stratum after having been in the developing village stratum. The findings of the literature study show that clustering methods can be used to perform data mining analysis.

The dataset was obtained from the West Java Provincial government. The initial data resulting from the Selection on the village portal page has 6 attributes, namely as follows: District Name, Sub-district Name, Healthy Family Index, Energy Consumption Intensity, Environmental Health Inspection, and Village Development Index. The dataset has 930 records.

Selection of data so that it is in accordance with the research needs and the application of KDD planned in the research methodology. Attributes obtained from the West Java provincial government have been considered not to be used all in the data processing process. So that the attribute selection process is needed. The Selection process is done with the rapid miner application using the Select Attributes operator. Figure 4 shows the Selection view in the rapid miner application.

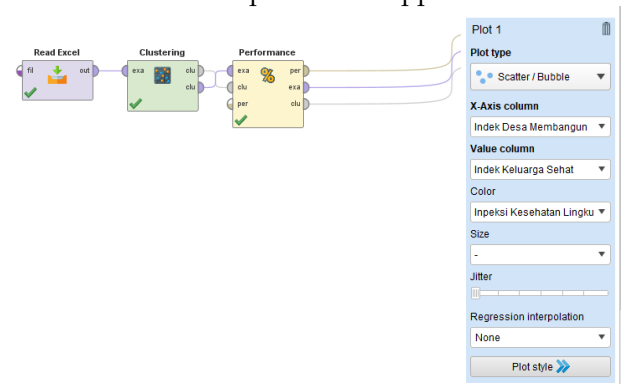


Figure 4. Select attribute design

Figure 4 displays the selection process of attributes to be used. The attributes selected in data processing

consist of 6 attributes, namely: District, sub-district, healthy family index, energy consumption intensity, environmental health inspection, and village development index. The attributes of district and sub-district names are used as type id so that in data processing they will not be taken into account, because these attributes are the object of research to determine the results of clustering.

Preprocessing is repairing damaged data to make it consistent into data that can be processed, as well as reducing data that is difficult to understand. However, the data used in this research turned out to have no errors. So this process can proceed to the transformation stage. Districts and sub-districts are transformed into consecutive numbers from each different district and sub-district.

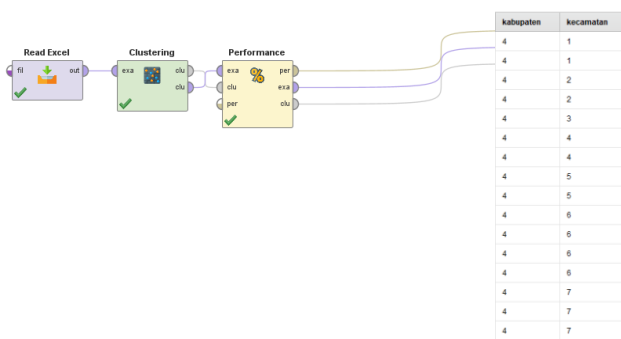


Figure 5. Data Transformation

Figure 6 shows the design of the data transformation process with the process results known by their respective numerical values. The clustering method with K-Means algorithm is used to provide all transformed data for mining. The dataset that has passed the preprocessing and transformation stages is now ready for the next stage, the mining process. Currently, there are 930 valid records with 6 attributes, resulting in the cluster model in Figure 6.

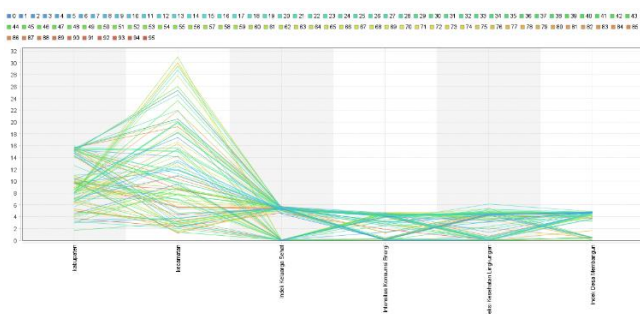


Figure 6. Cluster Model of Developing Villages in West Java

Figure 6 displays the membership of each Cluster. Cluster 0: 10 villages Cluster 1: 8 villages, Cluster 2: 12 villages, Cluster 3: 8 villages, Cluster 4: 19 villages, Cluster 5: 20 villages, Cluster 6: 7 villages, Cluster 7: 12 villages, Cluster 8: 15 villages, Cluster 9: 16 villages,

Cluster 10: 7 villages, Cluster 11: 6 villages, Cluster 12: 17 villages, Cluster 13: 9 villages, Cluster 14: 11 villages, Cluster 15: 14 villages, Cluster 16: 5 villages, Cluster 17: 17 villages, Cluster 18: 4 villages, Cluster 19: 4 villages, Cluster 20: 5 villages, Cluster 21: 3 villages, Cluster 22: 18 villages, Cluster 23: 11 villages, Cluster 24: 6 villages, Cluster 25: 3 villages, Cluster 26: 12 villages, Cluster 27: 8 villages, Cluster 28: 6 villages, Cluster 29: 6 villages, Cluster 30: 11 villages, Cluster 31: 3 villages, Cluster 32: 10 villages, Cluster 33: 5 villages, Cluster 34: 10 villages, Cluster 35: 15 villages, Cluster 36: 11 villages, Cluster 37: 17 villages, Cluster 38: 6 villages, Cluster 39: 17 villages, Cluster 40: 7 villages, Cluster 41: 6 villages, Cluster 42: 5 villages, Cluster 43: 3 villages, Cluster 44: 26 villages, Cluster 45: 16 villages, Cluster 46: 9 villages, Cluster 47: 3 villages, Cluster 48: 10 villages, Cluster 49: 8 villages, Cluster 50: 8 villages, Cluster 51: 23 villages, Cluster 52: 7 villages, Cluster 53: 2 villages, Cluster 54: 7 villages, Cluster 55: 1 village, Cluster 56: 24 villages, Cluster 57: 9 villages, Cluster 58: 1 village, Cluster 59: 9 villages, Cluster 60: 6 villages, Cluster 61: 8 villages, Cluster 62: 5 villages, Cluster 63: 7 villages, Cluster 64: 1 village, Cluster 65: 2 villages, Cluster 66: 10 villages, Cluster 67: 16 villages, Cluster 68: 21 villages, Cluster 69: 19 villages, Cluster 70: 10 villages, Cluster 71: 10 villages, Cluster 72: 10 villages, Cluster 73: 1 village, Cluster 74: 4 villages, Cluster 75: 8 villages, Cluster 76: 2 villages, Cluster 77: 18 villages, Cluster 78: 3 villages, Cluster 79: 30 villages, Cluster 80: 8 villages, Cluster 81: 7 villages, Cluster 82: 3 villages, Cluster 83: 3 villages, Cluster 84: 9 villages, Cluster 85: 3 villages, Cluster 86: 8 villages, Cluster 87: 20 villages, Cluster 88: 12 villages, Cluster 89: 11 villages, Cluster 90: 7 villages, Cluster 91: 4 villages, Cluster 92: 9 villages, Cluster 93: 13 villages, Cluster 94: 7 villages, Cluster 95: 27 villages with Total number of villages: 930



Figure 7. Davies Bouldin Index

The performance of an algorithm from the analysis results can be presented in various ways. This research presents the performance of the algorithm by showing the value of the Davies Bouldin Index. In Figure 7 the Davies Bouldin Index is generated with a value of -0.966 using a value of K = 96. Table 2 is a comparative value of the performance of the algorithm in processing data. Algorithms that have good performance also tend to be stable and responsive to various situations or different

types of input. By considering these factors, developers can select or optimise algorithms to achieve the best results in solving a problem or computational task.

Table 2. Performance comparison values

Nilai	Davies Bouldin Index
2	-1.035
3	-1.367
4	-1.307
5	-1.248
6	-1.246
7	-1.233
8	-1.404
9	-1.457
96	-0.966

Testing the performance of the algorithm in processing data was carried out 96 times. The first test was carried out by determining the value of $k = 2$ resulting in a Davies Bouldin Index value of -1.035. Then continue testing the performance of the algorithm by determining the value of $k = 3$ and so on until the value of $k = 96$. To find out the best performance value in data processing, the smallest Davies Bouldin Index value (close to 0) is selected. From the results of the Davies Bouldin Index value in Table 2, it can be compared that the smallest value of Davies Bouldin Index is at a value of $k = 96$ with a value of Davies Bouldin Index = -0.966 so it can be concluded that dividing the data into 96 clusters has a better accuracy value compared to other clusters. The results of the study found that forming the data into 96 clusters is in line with the research of (Xiaoqiong & Zhang, 2011). (Xiaoqiong & Zhang, 2020), (Santos & Pedrini, 2016), (Kakoudakis et al., 2017), (Varanasi & Tripathi, 2019), (Anantathanavit & Munlin, 2016). Some of these studies processed data into several clusters. As suggested in the study (Owsiński et al., 2017) it is necessary to analyse the number of clusters to obtain optimal clustering results. Therefore, the novelty of this research is testing by obtaining the Davies Bouldin Index value in each cluster test. Cluster testing was conducted 96 times starting from the value of $K = 2$ to the value of $K = 96$. The test results show that the number of clusters as many as 96 groups with a value of -0.966 is the best performance compared to testing using other clusters.

Conclusion

In this study the data totalled 930 records. After preprocessing, 930 records remain and can be analysed. The search results are the latest results of testing the K-Means algorithm in data processing by displaying the Davies Bouldin index value. The novelty of this research comes from several previous studies that have not tested the performance of the K-Means algorithm in data

processing and strengthened by research from (Zhang et al., 2019). (Zhang et al., 2019) in 2019 who conducted an analysis to determine the number of clusters. On research (Syoer & Wahyudin, 2021) the most optimal number of groups for grouping villages in East Kalimantan Province based on the algorithm, namely as many as 4 groups. Based on the test results, the K-means algorithm shows the best performance by dividing the data into 96 groups, resulting in a Davies Bouldin index value = -0.996. Based on the results of cluster division, cluster 0 has 10 villages, cluster 1 has 8 villages, cluster 2 has 12 villages. For this reason, this study recommends that the West Java provincial government needs to support and fulfil the needs of the village. Based on the results of the research that has been done, the author suggests continuing this research by re-analysing cluster identification so as to obtain a smaller Davies Bouldin index value (closer to 0).

Acknowledgments

The researcher would like to thank the Institute for Research and Community Service of Darwan Ali University for giving us permission to conduct research on the clustering of village development in West Java. We would also like to thank the West Java Provincial Government for being transparent about village data that supports this research so that it can run until it is completed. Then we would also like to thank fellow lecturers of Darwan Ali University who have provided moral support so that this research can be completed properly.

Author Contributions

Nurani Aisiyah Tanjung: preparation for writing - initial draft, results, discussion, methodology, conclusion, review, and editing; Nurahman: analysis and proofreading.

Funding

This research was independently funded by researches

Conflicts of Interest

Conflict of Interest The authors stated that they wanted to help the West Java provincial government determine which developing stratum villages could potentially enter the developed village stratum in the following year.

References

- Adinugroho, S., Wihandika, R. C., & Adikara, P. P. (2022). Newsgroup topic extraction using term-cluster weighting and Pillar K-Means clustering. *International Journal of Computers and Applications*, 44(4), 357-364. <https://doi.org/10.1080/1206212X.2020.1757246>
- Agarwal, P., & Mehta, S. (2016). Enhanced flower pollination algorithm on data clustering. *International Journal of Computers and Applications*, 38(2-3), 144-155. <https://doi.org/10.1080/1206212X.2016.1224401>

- Alhapizi, M. R., Nasir, M., & Effendy, I. (2020). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang. *Journal of Software Engineering Ampera*, 1(1), 1-14. <https://doi.org/10.51519/journalsea.v1i1.10>
- Anantathanavit, M., & Munlin, M. (2016). Using K-means radius particle swarm optimization for the travelling salesman problem. *IETE Technical Review (Institution of Electronics and Telecommunication Engineers, India)*, 33(2), 172-180. <https://doi.org/10.1080/02564602.2015.1057770>
- Chen, J. H., Chang, Y. C., & Hung, W. L. (2020). A self-organizing clustering algorithm for functional data. *Communications in Statistics: Simulation and Computation*, 49(5), 1237-1263. <https://doi.org/10.1080/03610918.2018.1494280>
- DPM-Desa Provinsi Jawa Barat. (2023). PORTAL BERDESA JAWA BARAT. Portaldata.desa. <https://portaldata.desa.jabarprov.go.id/>
- Economou, P. (2023). A clustering algorithm for overlapping Gaussian mixtures. *Research in Statistics*, 1(1), 0. <https://doi.org/10.1080/27684520.2023.2242337>
- Feng, Z., & Zhang, J. (2022). Nonparametric K-means algorithm with applications in economic and functional data. *Communications in Statistics - Theory and Methods*, 51(2), 537-551. <https://doi.org/10.1080/03610926.2020.1752383>
- Filki, Y. (2022). Algoritma K-Means Clustering dalam Memprediksi Penerima Bantuan Langsung Tunai (BLT) Dana Desa. *Jurnal Informatika Ekonomi Bisnis*, 4, 166-171. <https://doi.org/10.37034/infeb.v4i4.166>
- Fuss, C. E., Berg, A. A., & Lindsay, J. B. (2016). DEM Fusion using a modified k-means clustering algorithm. *International Journal of Digital Earth*, 9(12), 1242-1255. <https://doi.org/10.1080/17538947.2016.1208685>
- Gunawan, G., Jufri, A. W., Sedijani, P., Hadiprayitno, G., & Bachtiar, I. (2019). Pelatihan Penulisan Pada Jurnal Internasional Bagi Dosen Dan Mahasiswa Magister Pendidikan Ipa Unram. *Jurnal Pendidikan Dan Pengabdian Masyarakat*, 2(4). <https://doi.org/10.29303/jppm.v2i4.1526>
- Kakoudakis, K., Behzadian, K., Farmani, R., & Butler, D. (2017). Pipeline failure prediction in water distribution networks using evolutionary polynomial regression combined with K-means clustering. *Urban Water Journal*, 14(7), 737-742. <https://doi.org/10.1080/1573062X.2016.1253755>
- Kamalha, E., Kiberu, J., Nibikora, I., Mwasiagi, J. I., & Omollo, E. (2018). Clustering and Classification of Cotton Lint Using Principle Component Analysis, Agglomerative Hierarchical Clustering, and K-Means Clustering. *Journal of Natural Fibers*, 15(3), 425-435. <https://doi.org/10.1080/15440478.2017.1340220>
- Kim, H. S., Kim, S. K., & Kang, L. S. (2021). BIM performance assessment system using a K-means clustering algorithm. *Journal of Asian Architecture and Building Engineering*, 20(1), 78-87. <https://doi.org/10.1080/13467581.2020.1800471>
- Kouhi Esfahani, R., Shahbazi, F., & Akbarzadeh, M. (2019). Three-phase classification of an uninterrupted traffic flow: a k-means clustering study. *Transportmetrica B*, 7(1), 546-558. <https://doi.org/10.1080/21680566.2018.1447409>
- Li, J., Zheng, A., Guo, W., Bandyopadhyay, N., Zhang, Y., & Wang, Q. (2023). Urban flood risk assessment based on DBSCAN and K-means clustering algorithm. *Geomatics, Natural Hazards and Risk*, 14(1). <https://doi.org/10.1080/19475705.2023.2250527>
- Liu, Y., & Chen, J. (2019). A K-means based firefly algorithm for localization in sensor networks. *International Journal of Parallel, Emergent and Distributed Systems*, 34(4), 364-379. <https://doi.org/10.1080/17445760.2018.1446211>
- Okta Jaya Harmaja, Hadirat Halawa, Wicarda Sandi Hulu, & Seriani Loi. (2023). Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Pulo Brayon. *Jurnal Sains Dan Teknologi, Volume 5 No. 1(1)*, 150-157.
- Owsiński, J. W., Opara, K., Stańczak, J., Kacprzyk, J., & Zadrożny, S. (2017). Reverse clustering: an outline for a concept and its use. *Toxicological and Environmental Chemistry*, 99(7-8), 1078-1095. <https://doi.org/10.1080/02772248.2017.1333614>
- Pattamaset, S., & Choi, J. S. (2020). Irrelevant data elimination based on a k-means clustering algorithm for efficient data aggregation and human activity classification in smart home sensor networks. *International Journal of Distributed Sensor Networks*, 16(6). <https://doi.org/10.1177/1550147720929828>
- Reena Roy, R., & Anandha Mala, G. S. (2023). An Improved K-Means Clustering for Segmentation of Pancreatic Tumor from CT Images. *IETE Journal of Research*, 69(7), 3966-3973. <https://doi.org/10.1080/03772063.2021.1944335>
- Santos, A. C. S., & Pedrini, H. (2016). Kombinasi pengelompokan k-means dan pemfilteran entropi untuk pemilihan band dan klasifikasi pada citra hiperspektral. 1161, 3005-3020.
- Sioyong, D. (2023). Implementasi data mining menggunakan algoritma k-means clustering penyakit pasien rawat jalan pada klinik dr . Atirah. 7(5), 1-8.

- Syoer, R. R., & Wahyudin, Y. (2021). Analisis Kelompok Dengan Algoritma Fuzzy Clustering (Studi Kasus Pengelompokan Desa Di Provinsi Kalimantan Timur). *BESTARI: Buletin Statistika Dan Aplikasi Terkini*, 1, 1–11. <https://bestari.bpskaltim.com/index.php/bestari-bpskaltim/article/view/1/2>
- Varanasi, J., & Tripathi, M. M. (2019). K-means clustering based photo voltaic power forecasting using artificial neural network, particle swarm optimization and support vector regression. *Journal of Information and Optimization Sciences*, 40(2), 309–328. <https://doi.org/10.1080/02522667.2019.1578091>
- Vats, S., & Sagar, B. B. (2019). Performance evaluation of K-means clustering on Hadoop infrastructure. *Journal of Discrete Mathematical Sciences and Cryptography*, 22(8), 1349–1363. <https://doi.org/10.1080/09720529.2019.1692444>
- Virgo, I., Defit, S., & Yuhandri, Y. (2020). Klasterisasi Tingkat Kehadiran Dosen Menggunakan Algoritma K-Means Clustering. *Jurnal Sistim Informasi Dan Teknologi*, 2, 23–28. <https://doi.org/10.37034/jsisfotek.v2i1.17>
- Xiaoqiong, W., & Zhang, Y. E. (2020). Image segmentation algorithm based on dynamic particle swarm optimization and K-means clustering. *International Journal of Computers and Applications*, 42(7), 649–654. <https://doi.org/10.1080/1206212X.2018.1521090>
- Yolanda, E. (2023). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Data Pasien Rehabilitasi Narkoba. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(1), 182–191. <https://doi.org/10.30865/klik.v4i1.1107>
- Zhang, S., Li, S., Hu, J., Xing, H., & Zhu, W. (2019). An iterative algorithm for optimal variable weighting in K-means clustering. *Communications in Statistics: Simulation and Computation*, 48(5), 1346–1365. <https://doi.org/10.1080/03610918.2017.1414244>