JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 69 – 76 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3554
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

COMPARATIVE ANALYSIS OF K-MEANS, X-MEANS AND K-MEDOIDS IN CLASSIFYING MARRIAGE CHOICED ADMIST QUARTER-LIFE CRISIS

Disya Nurul Ariza1[1], Rahayu Ningsih[1]*, Sri Muryani[1], Herlina Ferliyanti[1], Ahmad Jurnaidi Wahidin[1]
[1]Engineering and Informatics, Universitas Bina Sarana Informatika
email: *rahayu.ryh@bsi.ac.id

Abstract: Bekasi Regency, being one of the key cities in Indonesia, offers a suitable setting to study the intricacies of marriage decision-making during a quarter-life crisis. This study focuses on the application of clustering algorithms to categorize individuals based on their marriage choices. Data was collected from a questionnaire completed by 110 respondents from Bekasi Regency, specifically individuals aged 18 to 30 who are single, including 80 women and 30 men. Data analysis was conducted using the RapidMiner software to evaluate the effectiveness of three clustering algorithms K-Means, X-Means, and K-Medoids in categorizing marriage decision patterns among young people experiencing a Quarter Life Crisis in Bekasi Regency. Results indicate that each algorithm has its own strengths and limitations in handling Quarter Life Crisis data.The results of the analysis show that the K-medoids algorithm provides the best clustering results with the lowest DBI value of 0.195, followed by the X-Means algorithm with a value of 0.199 and K-Means with a value of 0.207. These results can help understand the pattern of marriage decisions in the Quarter Life Crisis phase and help provide insights for policymakers in Bekasi Regency to make more effective intervention programs.

Keywords: K-Means; K-Medoids; X-Means

Abstrak: Sebagai salah satu kota besar di Indonesia, Kabupaten Bekasi memberikan konteks yang tepat untuk mempelajari kompleksitas pengambilan keputusan pernikahan di tengah krisis seperempat usia. Penelitian ini berfokus pada pemanfaatan algoritma clustering untuk mengelompokkan individu berdasarkan pilihan pernikahan mereka. Data diambil dari kuesioner yang diisi oleh 110 responden di Kabupaten Bekasi, yang terdiri dari individu lajang berusia 18 hingga 30 tahun, yaitu 80 perempuan dan 30 laki-laki. Analisis data dilakukan dengan perangkat lunak RapidMiner untuk mengevaluasi efektivitas tiga algoritma pengelompokan—K-Means, X-Means, dan K-Medoids—dalam mengelompokkan pola keputusan pernikahan di kalangan pemuda yang menghadapi Quarter Life Crisis di Kabupaten Bekasi. Hasilnya menunjukkan bahwa setiap algoritma memiliki keunggulan dan kelemahannya masing-masing dalam memproses data Quarter Life Crisis. Hasil analisis menunjukkan bahwa algoritma K-medoids memberikan hasil clustering terbaik dengan nilai DBI terendah yaitu 0.195, diikuti oleh algoritma X-Means dengan nilai 0.199 dan K-Means dengan nilai 0.207. Hasil ini dapat membantu memahami pola keputusan menikah pada fase Quarter Life Crisis dan membantu memberikan wawasan bagi pembuat kebijakan di Kabupaten Bekasi membuat program intervensi yang lebih efektif.

Kata kunci: K-Means; K-Medoids; X-Means

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 69 – 76 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3554
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

The decision to get married is one of the important stages in a person's life. However, in recent years, a phenomenon known as *the Quarter life crisis* has emerged as a factor influencing this decision-making process in the younger generation. *Quarter life crisis* is a psychological condition that generally occurs between the ages of 18 and 30, where individuals face confusion, anxiety, and uncertainty regarding life goals, careers, and personal relationships. [1]

As one of Indonesia's major cities, Bekasi Regency provides an appropriate context to examine the complexities of marriage decision-making amid a quarter-life crisis. This research centers on utilizing clustering algorithms to classify individuals based on their marriage choices. The study employs three clustering techniques: K-Means, X-Means, and K-Medoids. K-Means, a well-known and efficient method, has limitations in selecting the ideal number of clusters. K-Medoids, on the other hand, is similar to K-Means but offers better handling of outliers by using medoids as the center of clusters. X-Means extends K-Means, allowing it to determine the optimal number of clusters automatically through certain criteria.

To evaluate the clustering performance of these three algorithms on marriage decision classification within a Quarter Life Crisis, the researcher utilized the Davies Bouldin Index (DBI). This index assesses clustering validity by considering the ratio between cluster distances and the spread of each cluster. A lower DBI value reflects superior clustering quality. Through DBI analysis, this study seeks to compare the effectiveness of K-Means, X-Means, and K-Medoids in categorizing marriage-related decisions in the Quarter Life Crisis phase among residents of Bekasi Regency. It is anticipated that this research will enhance understanding of marriage decision trends and help determine the most suitable clustering algorithm for this purpose.

The previous research was conducted at the North Sulawesi Provincial DPRD Secretariat to determine the quality of employee performance. This study used clustering methods, including k-means, k-medoids, and x-means, which were then processed with the help of RapidMiner [2]. The following research discusses the Utilizing K-Means and K-Medoids for Categorizing Country Data Based on Socio-Economic and Health Criteria, with the objective of identifying the countries most in need of assistance from Non-Governmental Organizations (NGOs) [3]. The following research uses 609 landslide events in West Java Province during the 2019 period, aiming to group landslide-prone areas in West Java Province. For the calculations, the researcher used the K-Means and K-Medoids methods for comparison [4]. Previous studies in the Bekasi District Court applied K-Means clustering to the analysis of divorce cases [5].

The research for the chosen final project topic also incorporated the use of K-Means with SVM [6]. The application of K-Means for medication planning is also practiced at RSUD Pekanbaru [7]. The X-Means method is also implemented at CV Mega Baja for customer satisfaction analysis [8]. To find the best laptops, data mining is applied with the Apriori algorithm for recommendation purposes [9]. At Puskesmas Pasir Jaya, the K-Means and Fuzzy algorithms are employed to assess

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 69 – 76 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3554
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

and map toddler nutritional status [10]. The K-Means technique, powered by RapidMiner, is used to analyze and map measles immunization among toddlers by province [11].

METHOD

In this research, three algorithms—K-Means, X-Means, and K-Medoids—were utilized by the researchers to assess their effectiveness in clustering marriage decisions among individuals in Bekasi Regency undergoing a Quarter Life Crisis.

The first method employed is K-Means, K-Means is a data analysis technique in data mining that enables unsupervised data modeling. This method organizes data into clusters by dividing it into distinct partitions. [5]. K-Means is also the simplest clustering method that groups data into k groups based on the centroid of each group [12]. However, the results of K-Means are greatly influenced by the parameter k and the initialization of the centroid. In general, K-Means initializes the centroid randomly [6]. According to Adiya & Desnelita, K-Means is a technique for clustering non-hierarchical data, allowing it to be separated into two or more distinct groups.

The steps of *the K-Means method* : First, Start by randomly selecting k points as the center of the cluster, Second, each data point finds its closest cluster through the smallest distance to the cluster center. After every point is allocated to its nearest cluster, new cluster structures are formed, third the last step includes recalculating the cluster center, with this process continuing until the center position stabilizes of the centroid (center of the cluster) does not change significantly.

The second method employed is X-Means, X-Means is algorithm addresses certain limitations found in the K-Means method. A primary flaw of K-Means is the extended computational time needed for users to manually specify the number of clusters, or k. As an unsupervised learning technique, the X-Means algorithm performs data clustering automatically, eliminating the need for prior knowledge about which cluster each data point belongs to. [8]

The third method employed is K-Medoids**, t**he K-Medoids algorithm is a partitioning technique for data clustering that organizes a set of n objects into clusters. This method designates a single object from the dataset as the representative for each cluster, aiming to reduce the overall dissimilarity between objects by carefully choosing a reference point within each cluster. [13], [14]

In data mining, valuable information is uncovered from large datasets through the use of statistical methods, mathematical techniques, and AI, yielding deeper understanding. [9]. Data mining is a set of techniques to find previously unknown knowledge in large databases. The patterns found can be used to help make a decision [15].

While there are various methods and techniques for conducting data mining, this process typically requires systematic steps to ensure that relevant data can be collected effectively.

The main functions of data mining include: Descriptive function: To dig deeper into the details of the data being studied and Predictive functions: To build models that can predict patterns or outcomes from observed data.

DBI helps identify the optimal cluster count after clustering is complete.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 69 – 76 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3554
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

The purpose of DBI is to increase inter-cluster separation while reducing intra-cluster distances. This method serves as a benchmark for evaluating and comparing the effectiveness of clustering algorithms. A lower DBI indicates superior clustering results. DBI is calculated by finding the average for each point, dividing the cluster variance by the distance between cluster centers.

Clustering is a data analysis technique aimed at uncovering patterns in previously unexplored data. Commonly used in data mining, this method helps to identify and group similar data into distinct "zones."

Clustering is an unsupervised data mining technique, also referred to as "unsupervised learning." Data is classified into clusters, with each cluster containing points that are more alike within the group than they are compared to points in other clusters. Clustering seeks to partition the entire dataset into groups with relative similarity, maximizing similarity within a group while minimizing it between groups. [16][17], [18].

RapidMiner is a robust open-source platform for data science and machine learning, offering a wide range of tools for data preparation, modeling, evaluation, and deployment. Designed for ease of use, RapidMiner enables users to build and test multiple models effortlessly, even without programming expertise [19]. The software also features a range of built-in operators, which serve as the foundational elements of workflows and cover every stage of the data mining process, including data cleaning, feature selection, and modeling. RapidMiner applies various descriptive and predictive techniques to deliver insights, empowering users to make well-informed decisions [20]. The research method was prepared systematically using data from the results of questionnaires that had been filled out by respondents who had experienced and had passed the *Quarter Life Crisis*. From the data collected, 110 valid respondents were obtained and then processed to have met the criteria of this study. In addition, to process the research, the necessary hardware and software, including the Rapidminer application, to ensure the research findings are precise and as effective as possible. The stages of this research flow are as follows: First, Data Collection Stage : In this study, the data source used is questionnaire data that has been filled in by respondents in the form of answers to questions about *the Quarter life crisis*. Second, Data Processing Stage : Data Cleaning/ Selection Third Data transformation : Clustering, Evaluation, Result Analysis.

RESULT AND DISCUSSION

The questionnaire data is processed and input into Miscrosoft Excel so that the data can be used for the research process, then the data is summarized in such a way that the data can be processed by *the clustering* method, so that it becomes a new dataset that can be processed using the rapidminer application to group marriage decisions in *the Quarter life crisis*. After that, do cleaning / Selection of questio nnaire data by removing inappropriate attributes and missing values.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 69 – 76 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3554
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Figure 1. Data Selection

Furthermore, transform the data from questionnaire questions, to numerical to facilitate data processing. With yes it is worth 2 and if it is not worth 1. For the attributes name, gender, age and status are not changed to numerical.

Modeling of each case was done using the RapidMiner application. RapidMiner can make it easier for users to make estimates in large amounts of data using the operators needed. In this study, each step in modeling is carried out with several considerations. RapidMiner configuration includes determining the best optimization and best problem solving. After the modeling is carried out, the results obtained can be used to draw conclusions from the data processing that has been carried out.

After importing the data and checking the data, then modeling the dataset. The algorithm used is *the k-means* algorithm and evaluation using *Davies Bouldin*

Figure 2. Design Display in RapidMiner

The visualization uses 3D Scatter where the X-Axis Column and Value Column with Name and Y Axis and Color with Cluster. In this experiment, a clustering process was carried out, to see the clusters generated on each algorithm in the clustering.

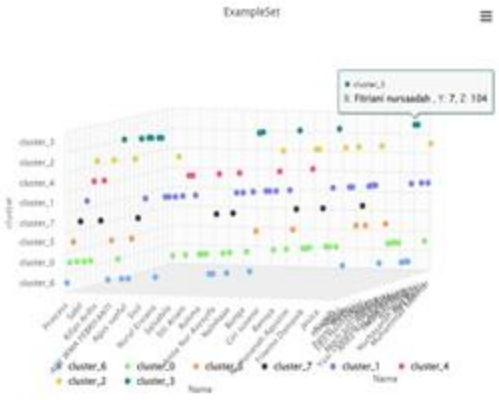

Figure 3. K-Means Algorithm Visualization Display

Figure 3 shows that the best clusteri ng in the K-Means algorithm is in cluster 3 where there are 12 people, the cluster with the highest number is in cluster 0 where there are 26 people and the lowest number of clustering is in clusters 4, 5, and 7 with 8 people.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 1, Desember 2024, hlm. 69 – 76 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i1.3554
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

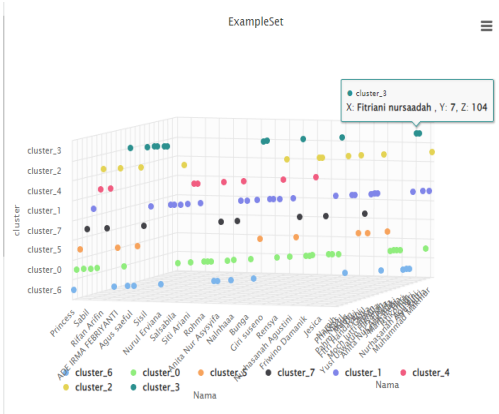

Figure 4. Visualization Display of X-Means Algorithm

Figure 4 shows that the best clustering in the X-Means algorithm is in cluster 7 where there are 3 people, the clustering with the highest number is in cluster 4 where there are 28 people and the clustering with the lowest number is in cluster 7 with 3 people.

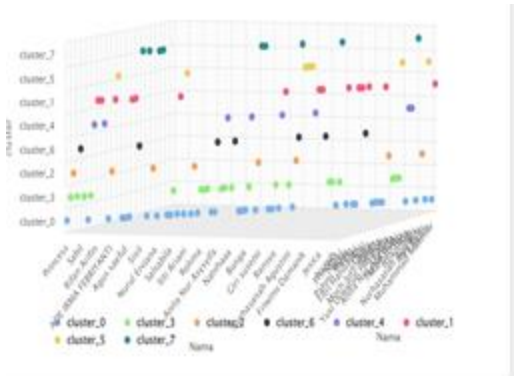

Figure 5. Visualization of K-Medoids Algorithm

Figure 5 shows that the best clustering in the K-Medoids algorithm is in cluster 7 where there are 9 people, the cluster with the highest number is in cluster 0 where there are 36 people and the clustering with the lowest number is in clusters 5 and 6 with 7 people.

The evaluation stage employs DBI by selecting the Cluster Distance Performance operator. Set the main criterion to Davies Bouldin, and choose the options to normalize and maximize, enabling measurement of data proximity within groups through standard deviation. Then, conduct experiments using the Davies Bouldin Index to compare the k-means, x-means, and k-medoids algorithms. This method aims to identify the most effective clustering algorithm.

In the last step, Davies Bouldin's index for each algorithm will be analyzed. This is done to find out the most suitable clustering method for each algorithm.

Tabel 1. *Davies Bouldin Index results*

Algoritma	Davies Bouldin Index Result
K-Means	0.207
X-Means	0.199
K-Medoids	0.196

CONCLUSION

The research concludes as follows: The K-Medoids algorithm proved to be the best method for grouping marriage decisions, achieving a Davies Bouldin Index value of 0.196, the lowest of all methods. According to K-Medoids visualization, the most accurate cluster is Cluster 7, with 9 respondents—1 male and 8 females aged 21 to 23; 7 are employed, and 2 are unemployed. All are unmarried and ready to decide on marriage during the Quarter Life Crisis. The X-Means algorithm ranked second, with a DBI of 0.199, slightly above K-Medoids by 0.003 and 0.008 above K-Means. In the X-Means visualization, Cluster 7 is the most relevant, containing

3 respondents—1 male and 2 females aged 21 to 23—who are employed, unmarried, have not experienced a Quarter Life Crisis, and are not yet prepared for marriage during this phase. For the K-Means algorithm, which achieved a DBI of 0.207, the most notable cluster is Cluster 3, with 12 respondents, 2 males and 10 females aged 21-23. In this group, 8 have jobs, 4 do not, all are unmarried, and 10 have faced a Quarter Life Crisis, while 2 have not. Among them, 3 are unprepared for marriage, while 9 are ready to decide on marriage during this period.

BIBLIOGRAPHY

[1] F. Akbar, "Apa Itu Quarter Life Crisis? Bagaimana Cara Kamu Menghadapinya?" [Online]. Available: https://satupersen.net/blog/quarter-life-crisis-bagaimana-kamu-menghadapinya

[2] G. B. Kaligis and S. Yulianti, "Analisa Perbandingan Algoritma K-Means, K-Medoids, dan x-Means Untuk Pengelompokkan Kinerja Pegawai (studi Kasus : Sekeretariat DPRD Provinsi Sulawesi Utara)," *IT-EXPLORE J. Penerapan Teknol. Inf. dan Komun.*, vol. 01, pp. 179–193, 2022.

[3] D. Hastari, F. Nurunnisa, S. Winanda, and D. D. Aprillia, "Application of K-Means and K-Medoids Algorithms for Grouping Country Data Based on Socio-Economic and Health Factors Penerapan Algoritma K-Means dan K-Medoids untuk Mengelompokkan Data Negara Berdasarkan Faktor Sosial-Ekonomi dan Kesehatan," pp. 274–281, 2023.

[4] M. Herviany, S. P. Delima, and T. Nurhidayah, "Comparison of K-Means and K-Medoids Algorithms for Grouping Landslide Prone Areas in West Java Province Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokkan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat," vol. 1, no. April, pp. 34–40, 2021.

[5] U. D. Rahayu *et al.*, "ANALISIS KASUS PERCERAIAN PADA PENGADILAN NEGERI BEKASI MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," *J. IKRAITH-INFORMATIKA*, vol. 6, no. 1, pp. 165–172, 2022.

[6] O. Somantri and S. Wiyono, "Metode K-Means untuk Optimasi Klasifikasi Tema Tugas Akhir Mahasiswa Menggunakan Support Vector Machine (SVM)," *JATI - J. Mhs. Tek. Inform.*, vol. 3, no. 1, pp. 34–45, 2016.

[7] M. H. Adiya and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," *J. Nas. Teknol. dan Sist. Inf.*, vol. 01, pp. 17–24, 2019.

[8] R. Adniana, D. Solihudin, and R. Narasati, "OPTIMASI ANALISIS DATA KEPUASAN PELANGGAN CV MEGA BAJA BINTARO DENGAN PENERAPAN ALGORITMA X-MEANS CLUSTERING," *JATI - J. Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 445–453, 2024.

[9] R. Mahmud and A. Hartanto, "PENERAPAN DATA MINING REKOMENDASI LAPTOP

MENGGUNAKAN ALGORITMA APRIORI," *JUISI*, vol. 06, no. 02, 2020.

[10] N. S. Fatonah and T. K. Pancarani, "ANALISA PERBANDINGAN ALGORITMA CLUSTERING UNTUK PEMETAAN STATUS GIZI BALITA DI PUSKESMAS PASIR JAYA," *KONVERGENSI*, vol. 18, no. 1, 2022.

[11] R. W. Sari, A. Wanto, and A. P. Windarto, "IMPLEMENTASI RAPIDMINER DENGAN METODE K-MEANS (STUDY KASUS : IMUNISASI CAMPAK PADA BALITA BERDASARKAN PROVINSI)," vol. 2, pp. 224–230, 2018.

[12] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J — Multidiscip. Sci. J.*, vol. 2, no. 16, pp. 227–235, 2019.

[13] H. Al Azies, F. A. Rohmatullah, H. B. Rochmanto, and D. Putri, "TOWARDS OPTIMIZATION : DATA-DRIVEN APPROACH K-MEDOIDS CLUSTERING ALGORITHM FOR REGIONAL EDUCATION QUALITY," vol. 12, no. 3, 2022.

[14] J. Eska, A. N. Sari, and A. P. Bisnis, "PENERAPAN METODE K-MEDOIDS TERHADAP ANGGARAN," vol. 4307, no. August, pp. 821–827, 2024.

[15] P. Bhatia, *Data Mining and Data Warehousing*. Cambridge University Press, 2019.

[16] D. T. Larose and C. D. Larose, *Discovering Knowledge In Data - An Introduction Of Data Mning*. Wiley, 2014.

[17] S. F. Intan, W. Elvira, S. Rahayu, and N. Nurfadilla, "Comparison of the K-Means and K-Medoids Algorithms for Grouping Student Expenditures Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Pengeluaran Mahasiswa," pp. 35–40, 2023.

[18] E. Tasia and M. Afdal, "Comparison Of K-Means And K-Medoid Algorithms For Clustering Of Flood-Prone Areas In Rokan Hilir District Perbandingan Algoritma K-Means Dan K-Medoids Untuk Clustering Daerah Rawan Banjir Di Kabupaten Rokan Hilir," vol. 3, no. 1, pp. 65–73, 2023.

[19] academy.rapidminer.com, "get-started-with-rapidminer-and-machine-learning," https://academy.rapidminer.com/. [Online]. Available: https://academy.rapidminer.com/learning-paths/get-started-with-rapidminer-and-machine-learning

[20] I. W. S. Wicaksana, L. Ambarwati, D. A. Baskoro, and D. A. C, *Belajar Data Mining Dengan Rapid Miner*. 2013.