International Journal of Engineering and Computer Science Applications (IJECSA)
Vol. 3, No. 2, September 2024, pp. 81∼88
ISSN: 2828-5611
DOI: doi.org/10.30812/ijecsa.v3i2.4601

Clustering Analysis of Umrah Pilgrim Data Based on the K-Medoid Method

Dias Nabila Huda, Anthony Anggrawan, Hairani Hairani
Universitas Bumigora, Mataram, Indonesia

Article Info

Article history:

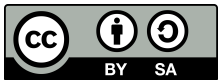
Received November 11, 2024
Revised November 17, 2024
Accepted November 22, 2024

Keywords:

Clustering Aalysis
Umrah Pilgrim
K-Medoid Method

ABSTRACT

The Umrah pilgrimage is becoming increasingly popular among Indonesians, with millions of participants yearly. This trend creates a need for service providers to understand the characteristics of pilgrims to improve service quality, marketing strategies, and competitiveness. Analyzing data on pilgrims helps service providers develop more effective strategies and tailor packages to match their needs, ensuring competitiveness in a growing market. This study aims to clusters Umrah pilgrims based on age, gender, district, and chosen package using the K-Medoid clustering method. This research uses the K-Medoid method for the reason that it is more resistant to noise and outliers compared to other clustering methods. The most centrally located point in the data set is called a "medoid," which is an object in a cluster that has the lowest difference to all other objects in the cluster. The results of this study are that the K-Medoid method successfully grouped pilgrims into three clusters: Cluster 1 with 63 members, Cluster 2 with 25 members, and Cluster 3 with 25 members. The findings indicate that the Milad Mastour package is preferred by older pilgrims, primarily from Mataram and West Lombok. The Arbain package is favored by younger pilgrims from the same regions, while adult pilgrims mostly choose the Regular package. The implication of this research is that it can provide insights for service providers to design more specific programs that align with the profiles of pilgrims based on age and district.



Corresponding Author:

Hairani Hairani, 087839793970
Universitas Bumigora, Mataram, Indonesia,
Email: hairani@universitasbumigora.ac.id

How to Cite: D. N. Huda, A. Anggrawan, and H. Hairani, "Clustering Analysis of Umrah Pilgrim Data Based on the K-Medoid Method", International Journal of Engineering and Computer Science Applications (IJECSA), vol. 3, no. 2, pp. 81-88, Sep. 2024. doi: 10.30812/ijecsa.v3i2.4601.

1. INTRODUCTION

Umrah is a form of religious tourism that is increasingly popular among Indonesians [1]. Each year, millions of pilgrims embark on this journey, creating a substantial demand for reliable and efficient Umrah travel services [2]. Service providers must understand the characteristics of pilgrims to improve their services and optimize revenue [3]. Analyzing Umrah pilgrim data provides valuable insights for developing more effective marketing strategies and services [4]. Furthermore, the growing number of Umrah service providers has intensified competition in the industry [5]. To remain competitive, service providers must understand and classify pilgrims based on their characteristics to offer more tailored packages [6].

A deeper understanding of Umrah pilgrim data can aid service providers in making better decisions [2]. By employing data analysis techniques such as clustering, providers can identify patterns and trends that might be overlooked through traditional data analysis methods [7] [8]. This allows providers to anticipate customer needs and design service packages that better align with customer preferences, enhancing customer retention and loyalty [9] [10]. While clustering approaches have been widely applied in various contexts, specific research on Umrah pilgrim data analysis remains limited. Some prior studies have used clustering methods for Umrah data analysis. For instance, a study [11] categorized Umrah pilgrims into three groups—highly preferred, preferred, and less preferred—based on attributes like gender, age, and region of origin, using the K-Means method. Study [12] also utilized the K-Means method to group Umrah travel data into three categories: highly popular, moderately popular, and less popular.

Study [13] employed the K-Means method to classify Umrah pilgrims into three groups: VIP, Executive, and Economy. Study [14] used the K-Nearest Neighbors (KNN) method to predict factors influencing Umrah pilgrims to cancel their departure. Study [15] identified patterns in hajj departures using the C4.5 algorithm. Study [16] categorized Umrah pilgrims into three groups—highly popular, popular, and less popular—based on attributes such as gender, age, and package preferences, using the K-Means method.

Previous studies show that using clustering in Umrah pilgrim data analysis remains limited. Moreover, previous research gaps often relied on the K-Means method, which only analyzed a few attributes. This study addresses these gaps by applying a different method and broader attributes. Specifically, the K-Medoid method is used for clustering, with attributes including age, gender, region, and Umrah package, which have not been previously explored. The selection of the K-Medoid method in this study is more resistant to noise and outliers than other clustering methods. The most centrally located point in the data set is called a ”medoid,” which is an object in a cluster that has the lowest difference to all other objects in the cluster. This study aims to group Umrah pilgrims based on these attributes using a clustering approach.These findings are expected to contribute to help Umrah service providers design more effective marketing strategies and improve service quality.

2. RESEARCH METHOD

The research stages, shown in Figure 1, begin with data gathering, followed by data processing, the implementation of the K-Medoid method, and result interpretation. The data used consists of Umrah pilgrim records from PT. Masy'aril Haram Tours and Travel for the years 2019 and 2022, comprising 113 instances and four attributes: gender, age, region, and package. The data processing stage involves transformation processes, such as encoding, which converts categorical data into numerical format for analysis. The next stage is implementing the K-Medoid method, a clustering technique similar to K-Means but with a key difference: K-Medoid uses medoids (actual data points) as cluster centers instead of centroids (the average of data points). This makes K-Medoid more robust to outliers, as medoids are real data points, whereas outliers can influence centroids [17, 18]. The steps of the K-Medoid method include [19, 20]: (1) Randomly selecting k medoids from the dataset; (2) Assigning each data point to the nearest medoid based on Euclidean distance; (3) Replacing medoids within clusters if other points serve as better medoids, reducing the total distance of points to the medoid; (4) Repeating steps 2 and 3 until medoids or clusters stabilize (convergence); (5) Determining the final clusters based on the resulting medoids, grouping each data point to the nearest medoid.

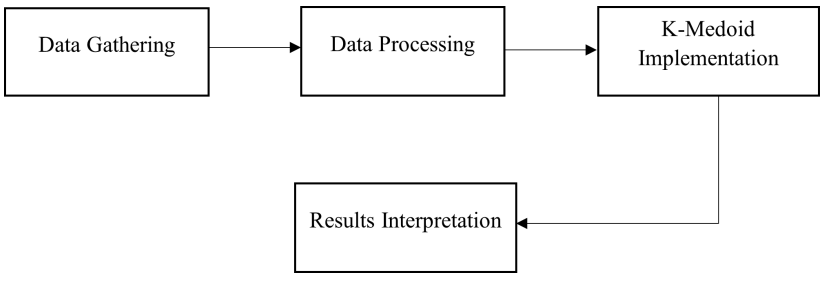


Figure 1. Research Flow

3. RESULT AND ANALYSIS

The initial data of Umrah pilgrims is presented in Table 1, including gender, age, region, and the selected Umrah package. This data consists of 113 pilgrims with various characteristics identified through these attributes. A transformation process was carried out to analyze this data using the K-Medoid clustering method. The transformation process is used to convert categorical data into a numerical format. Table 2 shows the transformation results for the gender attribute, where ”Male” is assigned the value 1 and ”Female” is the value 2. Table 3 presents the transformation for the region attribute, with each region assigned a unique value, such as ”Kota Mataram” valued at 1, ”Lombok Barat” valued at 2, ”Lombok Tengah” valued at 3, and ”Lombok Timur” valued at 4. Meanwhile, Table 4 shows the transformation for the Umrah package attribute, where each package type is assigned a specific value. After the transformation process, Table 5 displays the numerical data of the Umrah pilgrims, which is ready for the clustering process. This data was then used in the K-Medoid model to group the pilgrims based on certain patterns potentially related to gender, age, region, and package preferences. The clustering of Umrah pilgrims' data using K-Medoid was performed with the RapidMiner tool, as shown in Figure 2, and the clustering results are displayed in Table 6.

Table 1. Umrah Pilgrim Data

No	Gender	Age	City/Regency	Package
1	Male	46	Lombok Barat	Milad Mastour
2	Female	39	Lombok Barat	Milad Mastour
3	Male	12	Lombok Barat	Milad Mastour
4	Male	20	Lombok Barat	Milad Mastour
5	Female	8	Lombok Barat	Milad Mastour
...	...	...	...	...
109	Male	51	Lombok Timur	Regular Program
110	Female	31	Lombok Timur	Regular Program
111	Female	45	Lombok Timur	Regular Program
112	Female	25	Lombok Timur	Regular Program
113	Male	56	Lombok Timur	Regular Program

Table 2. Gender Transformation

No	Gender	Value
1	Male	1
2	Female	2

Table 3. City/Regency Transformation

No	City/Regency	Nilai
1	Kota Mataram	1
2	Lombok Barat	2
3	Lombok Tengah	3
4	Lombok Timur	4

Table 4. Program Package Transformation

No	Paket	Nilai
1	Milad Mastour	1
2	Arbain Program	2
3	Regular Program	3

Table 5. Umrah Pilgrim Data After Transformation

No	Gender	Age	City/Regency	Paket
1	1	46	2	1
2	2	39	2	1

No	Gender	Age	City/Regency	Paket
3	1	12	2	1
4	1	20	2	1
5	2	8	2	1
...	...	...	...	...
109	1	51	1	3
110	2	31	1	3
111	2	45	1	3
112	2	25	1	3
113	1	56	1	3

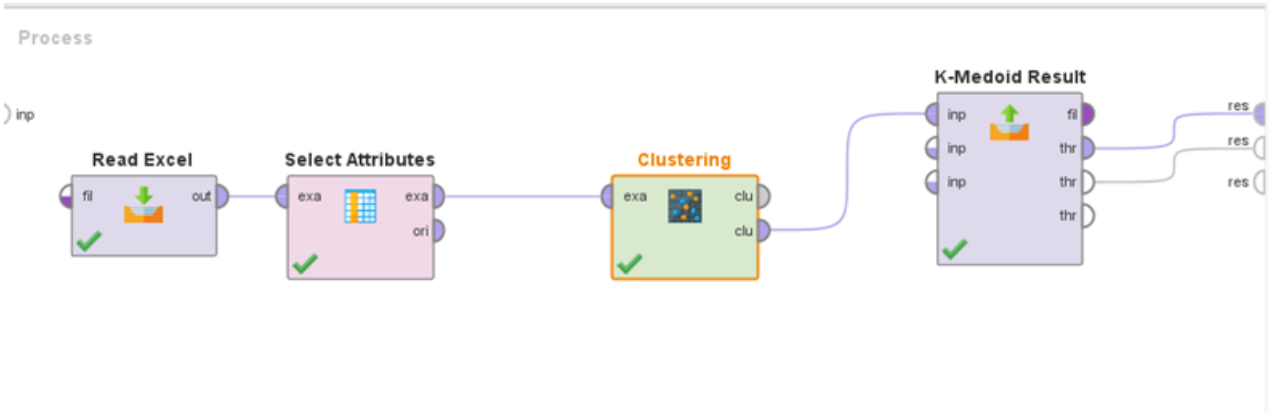


Figure 2. K-Medoid Model on Umrah Pilgrim Data Grouping

Table 6. Umrah Pilgrims Data Cluster Results

No	Gender	Age	City/Regency	Package	Cluster
1	1	46	2	1	3
2	2	39	2	1	3
3	1	12	2	1	2
4	1	20	2	1	2
5	2	8	2	1	1
...	...	...	...	...	...
109	1	51	1	3	1
110	2	31	1	3	2
111	2	45	1	3	3
112	2	25	1	3	2
113	1	56	1	3	1

The K-Medoid method successfully grouped the Umrah pilgrim data into three clusters, as shown in Figure 3: Cluster 1 with 63 people, Cluster 2 with 25 people, and Cluster 3 with 25 people. Based on the age distribution in Table 6, Cluster 1 is dominated by pilgrims aged 51 to 87 years, with an average age of 60; Cluster 2 is dominated by pilgrims aged 8 to 35 years, with an average age of 25; and Cluster 3 is dominated by pilgrims aged 36 to 50 years, with an average age of 42. Based on the gender distribution in Table 7, Cluster 1 consists of 32 women and 31 men, Cluster 2 is dominated by 15 women and 10 men, while Cluster 3 consists of 13 women and 12 men.

Based on the distribution of selected Umrah packages in Table 10, Cluster 1 consists of 39 individuals who chose the Milad Mastour package, 14 who selected the Arbain Program, and 10 who opted for the Regular Program. Cluster 2 includes 18 individuals who chose the Milad Mastour package, 1 who selected the Arbain Program, and 6 who opted for the Regular Program. Cluster 3 comprises 16 individuals who chose the Milad Mastour package, 2 who selected the Arbain Program, and 7 who opted for the Regular Program.

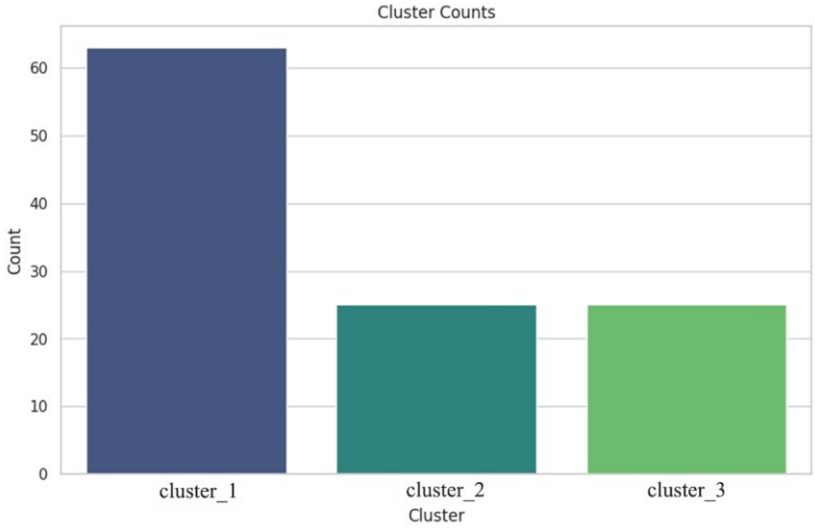


Figure 3. Umrah Pilgrims Clustering Results with K-Medoid

Table 7. Age Distribution by Cluster

Cluster	Total	Mean	Min	Max
1	63	60	51	87
2	25	25	8	35
3	25	42	36	50

Table 8. Gender Distribution by Cluster

Cluster	1 (Male)	2 (Female)
1	31	32
2	10	15
3	12	13

Table 9. City/Regency Distribution by Cluster

Cluster	1 (Kota Mataram)	2 (Lombok Barat)	3 (Lombok Tengah)	4 (Lombok Timur)
1	28	21	11	3
2	13	12	0	0
3	11	11	3	0

Table 10. Distribusi Paket Berdasarkan Cluster

Cluster	1 (Milad Mastour)	2 (Arbain Program)	3 (Regular Program)
1	39	14	10
2	18	1	6
3	16	2	7

The findings of this study indicate that the K-Medoid method effectively clusters Umrah package preferences. The Milad Mastour package is preferred by older individuals, particularly from the Mataram and Lombok Barat regions. The Arbain package is

mostly chosen by younger individuals, with most coming from the same regions, Mataram and Lombok Barat. Meanwhile, regular packages are popular among adults, mostly from Mataram and Lombok Barat.

Based on these findings, travel providers can consider package preferences according to the age profile and region of origin of the pilgrims, which aligns with the research by [21]. Considering package preferences based on age and regional profiles can assist providers in designing more tailored Umrah programs that meet the specific needs of pilgrims in each segment, thereby improving satisfaction and service efficiency.

4. CONCLUSION

This study shows that the K-Medoid method successfully grouped Umrah pilgrims' data into three clusters: cluster 1 with 63 individuals, cluster 2 with 25 individuals, and cluster 3 with 25 individuals. The analysis results reveal that package preferences vary according to the pilgrims' age and regional profiles (districts). Elderly pilgrims predominantly choose the Milad Mastour package from Mataram and Lombok Barat, the Arbain package by younger pilgrims from the same regions, and the Regular package by adult pilgrims. These findings assist Umrah providers in designing programs more suited to the pilgrims' profiles. A recommendation for future research is to compare the effectiveness of K-Medoid with other clustering methods, such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN), in grouping pilgrims based on their Umrah package preferences.

5. DECLARATIONS

AUTHOR CONTIBUTION

All authors contributed to the writing of this article

FUNDING STATEMENT

-

COMPETING INTEREST

The author declares no conflict of interest in this article.

REFERENCES

[1] D. Nurohman, N. R. Sari, R. Himmati, and R. Mas'ud, "Marketing Strategy of Umrah Travel Agent and Religious Attitude in Indonesia," *International Journal of Scientific Research and Management*, vol. 11, no. 02, pp. 4528–4529, 2023.

[2] M. Aulia, Y. Affandi, and A. Rozaq, "Tren dan Tantangan dalam Manajemen Industri Haji dan Umroh : Perspektif Global dan Lokal," *Jurnal Manajemen Pendidikan dan Ilmu Sosial*, vol. 5, no. 4, pp. 1224–1233, 2024.

[3] S. F. Rahman, M. Saggaff Shihab, and H. Hammad, "Membangun Kepuasan Pelanggan Jamaah Umroh Melalui Aktivitas Komunikasi Pemasaran dan Persepsi Harga," *Jurnal Administrasi dan Manajemen*, vol. 13, no. 3, pp. 294–306, 2023.

[4] D. K. Putri and A. Rozaq, "Studi Komparatif Pemasaran Paket Umroh dalam Meningkatkan Jumlah Jamaah pada PT . Mahabbah Wisata Semarang dan PT . Arrahmah Kendal," *Al Qalam: Jurnal Ilmah Keagamaan dan Kemasyarakatan*, vol. 18, no. 4, pp. 2339–2351, 2024.

[5] V. N. A. S. Romli, "Strategi Pemasaran Paket Ibadah Haji Dan Umroh Dalam Merekrut Jamaah Di PT. Sela Express Tour Kota Pekanbaru," *Jurnal Riset Mahasiswa Dakwah dan Komunikasi*, vol. 1, no. 9, pp. 1689–1699, 2019.

[6] I. Zulfikar and A. Palewa, "Analisis Strategi Pemasaran Dalam Rangka Meningkatkan Kualitas Pelayanan Terhadap Jamaah Umroh Pada PT. Daanish Mika Salsa Di Bandar Lampung," *Jurnal Manajemen Mandiri Saburai (JMMS)*, vol. 3, no. 01, pp. 1–8, 2020.

[7] M. Chaudhry, I. Shafi, M. Mahnoor, D. L. R. Vargas, E. B. Thompson, and I. Ashraf, "A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective," *Symmetry*, vol. 15, no. 9, pp. 1–44, 2023.

[8] D. Susilowati, H. Hairani, I. P. Lestari, K. Marzuki, and L. Z. A. Mardedi, "Segmentasi lokasi promosi penerimaan mahasiswa baru menggunakan metode rfm dan k-means clustering," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 2, pp. 275–282, 2022.

[9] B. Artha, I. Zahara, and N. P. Sari, "Customer Retention : A Literature Review," *Social Science Studies*, vol. 2, no. 1, pp. 30–45, 2022.

[10] H. Hairani and J. X. Guterres, "Exploring customer purchasing patterns: A study utilizing fp-growth algorithm on supermarket transaction data," *International Journal of Engineering and Computer Science Applications (IJECSA)*, vol. 3, no. 1, pp. 33–42, 2024.

[11] T. Hidayat, "Klasifikasi Data Jamaah Umroh Menggunakan Metode K-Means Clustering," *Jurnal Sistim Informasi dan Teknologi*, vol. 4, no. 1, pp. 19–24, 2022.

[12] M. Hermansyah, R. A. Hamdan, F. Sidik, and A. Wibowo, "Klasterisasi Data Travel Umroh di Marketplace Umroh.com Menggunakan Metode K-Means," *Jurnal Ilmu Komputer*, vol. 13, no. 2, p. 8, 2020.

[13] M. Yunitasari, T. Maharani, and B. Hikmahwan, "Implementasi Metode K-Means Untuk Pengelompokan Data Jamaah," *Kumpulan jurnaL Ilmu Komputer (KLIK)*, vol. 09, no. 1, pp. 1–9, 2022.

[14] A. Wijaya Kusuma and D. Mahdiana, "Development of Data Warehouse To Predicate the Regarding of Umroh Congregations Using the Manearest Neighbour Algorithm (Case Study Pt. Bahana Sukses Sejahtera)," *Jurnal Teknik Informatika (JUTIF)*, vol. 3, no. 4, pp. 1007–1012, 2022.

[15] F. A. Sianturi, M. Kumari, and E. Laian, "Implementasi Algortima C4.5 Menentukan Pola Berangkatan Jamaah Haji," *Jurnal Sains dan Teknologi*, vol. 3, no. 2, pp. 66–70, 2022.

[16] M. Iqbal, "Klasterisasi Data Jamaah Umroh Pada Auliya Tour & Travel Menggunakan Metode K-Means Clustering," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 5, no. 2, pp. 97–104, 2019.

[17] N. Rizqia and P. Ratnasari, "Comparative Study of k-Mean , k-Medoid, and Hierarchical Clustering," *Indonesia Journal of Life Sciences*, vol. 5, no. 2, pp. 9–20, 2023.

[18] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient," *Journal of Applied Informatics and Computing (JAIC)*, vol. 5, no. 2, pp. 109–116, 2021.

[19] H. L. Siregar, M. Zarlis, and S. Efendi, "Cluster Analysis using K-Means and K-Medoids Methods for Data Clustering of Amil Zakat Institutions Donor," *Jurnal Media Informatika Budidarma*, vol. 7, no. 2, p. 668, 2023.

[20] N. Mirantika, T. S. Syamfithriani, and R. Trisudarmo, "Implementasi Algoritma K-Medoids Clustering Untuk Menentukan Segmentasi Pelanggan," *Jurnal Nuansa Informatika*, vol. 17, no. 1, pp. 2614–5405, 2023.

[21] T. Hassan, M. Carvache-Franco, O. Carvache-Franco, and W. Carvache-Franco, "Sociodemographic relationships of motivations, satisfaction, and loyalty in religious tourism: A study of the pilgrimage to the city Mecca." *PloS one*, vol. 18, no. 3, pp. 1–17, 2023.

[This page intentionally left blank.]