Al-Hijr: Journal of Adulearn World, 3(4) - Dec 2024 437-445

The Use of Clustering Techniques to Enhance the Accuracy of Islamic Religious Knowledge Assessment

Muh. Nurtanzis Sutoyo [1], Nasruddin [2], Hasan Djidu [3]
[1] *Universitas Sembilanbelas November Kolaka, Indonesia*
[2] *Universitas Negeri Surabaya, Indonesia*
[3] *Universitas Negeri Yogyakarta, Indonesia*

Corresponding Author: Muh. Nurtanzis Sutoyo, E-mail: mns.usn21@gmail.com

Received: Sep 24, 2025	Revised: Feb 28, 2025	Accepted: Feb 28, 2025	Online: Feb 28, 2025

ABSTRACT
This study aims to enhance the accuracy of Islamic religious knowledge assessment among Madrasah students in Kolaka, Indonesia, by applying the K-Means clustering technique. The background of this research is based on the variation in students' understanding of two important aspects: Fiqh and Aqidah Akhlak. The method used includes data collection from 181 students, followed by descriptive analysis and determination of the number of clusters using the Elbow Method. The results show that the optimal number of clusters is three, representing the categories of High, Medium, and Low. After applying K-Means, a Silhouette Coefficient score of 0.548 was obtained, indicating that the student clustering was effective and valid. These findings provide insights for educators to design learning strategies that are more aligned with the students' needs and highlight the importance of using data analysis techniques in education. Thus, this study contributes to the development of more accurate assessment methods in the context of religious education.

Keywords: *Education Assessment, K-Means, Silhouette Coefficient*

Journal Homepage https://ejournal.staialhikmahpariangan.ac.id/Journal/index.php/alhjr/

How to cite: Sutoyo, N, M., Nasruddin, Nasruddin & Djidu, H. (2024). The Use of Clustering Techniques to Enhance the Accuracy of Islamic Religious Knowledge Assessment. *Al-Hijr: Journal of Adulearn World*, 3(4), 437-445. https://doi.org/10.55849/alhijr.v3i4.704
Published by: Sekolah Tinggi Agama Islam Al-Hikmah Pariangan Batusangkar

INTRODUCTION

Educational assessment is a key element in the education system, especially in the context of religious knowledge. Islamic religious knowledge, which encompasses various aspects such as Fiqh and Aqidah Akhlak, is crucial for every student to understand. Accurate assessment can provide a clear picture of students' understanding of the subject matter. However, the challenge in conducting assessments often lies in the complexity of the material and the heterogeneity of students' abilities. Therefore, a more systematic and effective approach is needed to evaluate Islamic religious knowledge, one of which is through the use of clustering techniques.

Clustering techniques, particularly the K-Means method, are data analysis methods that can help group data based on similar characteristics. In the context of assessing Islamic religious knowledge, K-Means can be used to cluster students based on their assessment results into three categories: Fiqh (X1) and Aqidah Akhlak (X2). By utilizing this technique, it is expected that the assessment will become more accurate, as it can identify patterns present in the exam results data.

Clustering is a technique in data analysis, especially in multivariate data processing, used to group objects or data with similar characteristics into several clusters (Faizan et al., 2020). Clustering is used to discover structures or patterns in unlabeled data sets and is often applied in exploratory data analysis. The main goal of clustering is to ensure that data within the same cluster have a high degree of similarity, while data from different clusters exhibit significant differences (Oyelade et al., 2019). Clustering is widely used in various fields such as computer science, biology, marketing, psychology, medicine, and many others.

K-Means is a clustering technique that falls under the category of unsupervised learning. Data is grouped into several clusters based on similar characteristics (Das, Kayal & Maiti, 2023). In subsequent research, the K-Means algorithm was used to classify COVID-19 cases in the Lampung Province. In this study, several methods can be employed to use data mining based on geographic zoning to better understand and manage the spread of COVID-19 (Nabila et al., 2021). K-Means aims to maximize the variation between clusters and minimize the variation within each cluster as much as possible (Yang & Hussain, 2023). K-Means is widely used in various applications such as market segmentation, computer organization, document clustering, satellite imagery, and many more, where understanding groups or hidden patterns in the data is highly beneficial.

Several previous studies using the K-Means Algorithm include examining the use of K-Means Clustering in the marketing strategy of a property company in Sidoarjo. The focus was on mapping the customer distribution to effectively identify customer segments. The analysis results showed that customer clustering based on geographic location can optimize marketing resources with more targeted efforts (Afifah, Nurcahyawati & Hananto, 2023). Furthermore, the K-Means Clustering method is used to classify stocks in the energy industry into groups based on their main features. This method is effective in helping investors select stocks that align with their desired characteristics, based on factors such as profitability, debt ratio, and company stability (Srimurdianti et al., 2023). Subsequently, the K-Means clustering method was used to group eucalyptus land in Cilacap Regency based on these factors (Maharrani, Abda'u & Hastuti, 2022).

Subsequently, the research explores file system security by analyzing user activity logs through data mining and the K-Means Clustering technique (Pamuji, 2022). Then, the research used the K-Means clustering algorithm to classify the impact of earthquakes in Java Island. This data was collected from the National Disaster Management Agency (BNPB) from 2012 to 2021 (Wahyu & Rushendra, 2022). The research used the K-Means clustering algorithm to group sales data of rattan furniture at PT. Hymsa Indotraco. The main objective was to identify the most frequently ordered products, rarely ordered

products, and highly frequently ordered products to assist management in planning raw material inventory (Febianda, Ratnawati & Rahayudi, 2020). The research discusses the use of the K-Means algorithm to cluster high-achieving students who apply for scholarships (Sovia, Mandala & Mardhiah, 2020), YouTube channel clustering (Badruttamam, Sudarno & Maruddani, 2020), Village potential clustering (Muningsih et al., 2021), Road development clustering (Kurniadi et al., 2023), and text mining tweets on the Blibli account (Fani, Santoso & Suparti, 2021).

On the other hand, the assessment of Islamic religious knowledge in schools often still relies on traditional methods that do not account for variations in students' understanding. Inaccurate assessment methods can lead to errors in decision-making, both for the students themselves and for educational institutions. By using clustering techniques, assessments can be conducted more objectively and accurately, providing more useful information for curriculum development and teaching strategies.

The objective of this research is to analyze and implement clustering techniques, particularly K-Means, in the assessment of Islamic religious knowledge. This study aims to improve the accuracy of the assessments given to students and to explore patterns in the understanding of subject matter that may not be detected through traditional assessment methods. Through this analysis, it is expected to provide recommendations for educators and educational institutions in formulating more effective and responsive teaching strategies tailored to students' needs.

The novelty of this research lies in the application of the K-Means technique in the context of Islamic religious knowledge assessment, which has rarely been done in previous studies. By integrating data analysis and education, this research is expected to make a significant contribution to the field of Islamic education and serve as a reference for further research on more innovative and effective evaluation techniques.

RESEARCH METHODOLOGY

This research employs a quantitative approach with an experimental research design. The primary objective is to analyze and apply the K-Means clustering technique in the assessment of Islamic religious knowledge in schools. This study will involve students taking exams in two categories: Fiqh (X1) and Aqidah Akhlak (X2).

The data used in this research are the assessment results of students in the Islamic religious education subject. The assessment is conducted through questions specifically designed for each category. Each student will receive a score based on the correct answers from the total questions given. The data collected will consist of two main variables: Fiqh (X1) and Aqidah Akhlak (X2).

Before using the K-Means algorithm, the ideal number of clusters must be determined. The Elbow Method is used for this purpose. In the Elbow Method, the SSE (Sum of Square Error) value for each cluster can be calculated. The SSE value decreases as the K value, or the number of clusters, increases (Guntara & Lutfi, 2023; Hartanti, 2020; Orisa, 2022).

Next, the K-Means Algorithm is applied to the data after determining the number of clusters. This algorithm will group Islamic religious knowledge into categories based on the similarity of attributes between data points. The non-hierarchical method using the K-Means algorithm divides the data into one or more clusters. These clusters consist of data with similar attributes and data with differing attributes. Due to its ease of use, this algorithm is one of the most popular. However, a drawback of this algorithm is that it is highly sensitive to cluster initialization (Sopyan, Lesmana & Juliane, 2022).

The clustering results will be evaluated using internal and external metrics, namely the Silhouette Score. The quality and strength of the clusters, as well as how well an object is placed within a cluster, can be measured using the Silhouette Coefficient (Afifah et al., 2023; Guntara & Lutfi, 2023; Nugroho et al., 2022; Paembonan & Abduh, 2021).

Meanwhile, to assess the quality of the Silhouette Coefficient, it can be observed in Table 1 below.

Table 1. Interpretation of Silhouette Coefficient Values

Silhouette Coefficient Value	Interpretation
0.71 – 1.00	Strong clustering structure
0.51 – 0.70	Moderate clustering structure
0.26 – 0.50	Weak clustering structure
≤ 0.25	No substantial clustering

RESULT AND DISCUSSION

The research data was obtained from students at several Madrasahs located in Kolaka, Indonesia. A total of 181 students participated in this study. The assessment of Islamic religious knowledge collected focused on two main aspects: Fiqh (X1) and Aqidah Akhlak (X2). The data was collected through exams consisting of multiple-choice questions designed for each category, aimed at measuring the students' understanding of the material taught.

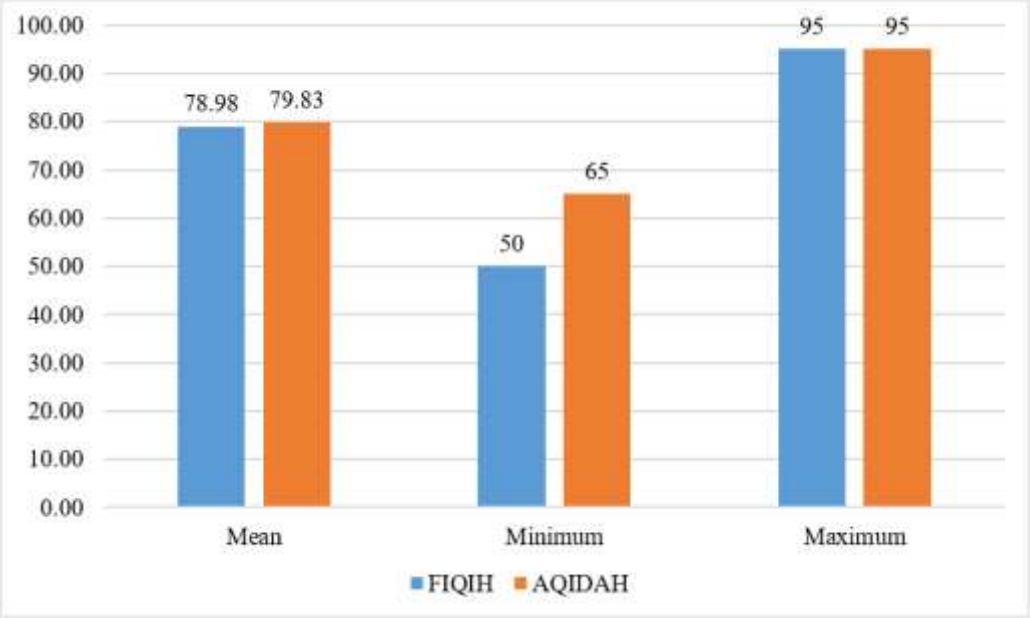
After data collection, descriptive analysis was conducted to obtain a general overview of the students' assessment results. Table 1 below presents the statistical description of the two analyzed variables.

Table 2. Statistical Description of Islamic Religious Knowledge Assessment Results

Statistics			
		FIQIH	AQIDAH
N	Valid	181	181
	Missing	0	0
Mean		78.98	79.83
Median		80.00	80.00
Std. Deviation		9.275	9.005
Variance		86.033	81.083
Minimum		50	65
Maximum		95	95

Figure 1 below shows a comparison chart of the maximum, minimum, and average scores for the two aspects of Islamic religious knowledge measured.

Figure 1. Score Comparison Chart

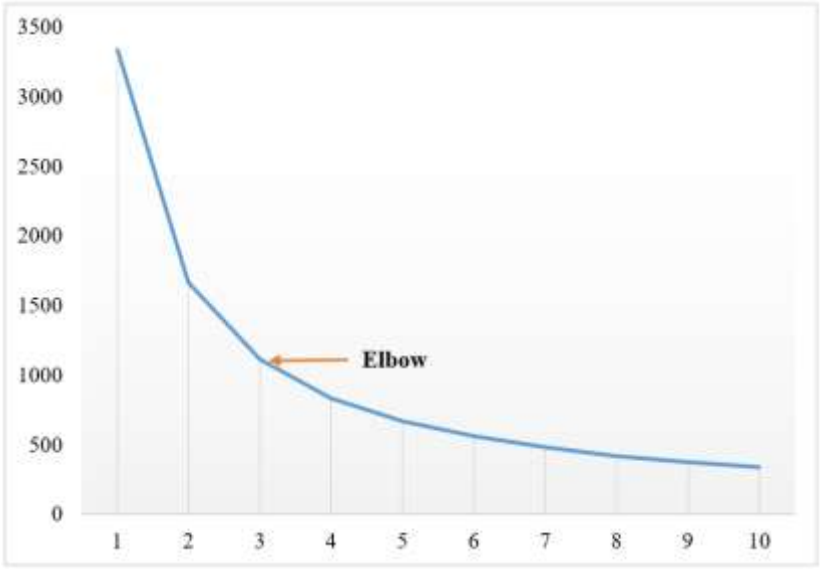


The statistical description shows that the average score for students' Fiqh knowledge (X1) is 78.98 with a standard deviation of 9.275, indicating that most students scored around this average. The lowest score in the Fiqh aspect was 50, while the highest reached 95. Meanwhile, in the Aqidah Akhlak variable (X2), the average student score was 79.83 with a standard deviation of 9.005. This indicates that students' understanding in the Aqidah Akhlak aspect tends to be better compared to Fiqh, with the lowest score being 65 and the highest 95.

After data collection and descriptive analysis, the next step in this research is to determine the optimal number of clusters for K-Means analysis. The method used is the Elbow Method, which aims to identify the 'elbow' point in the graph that shows the relationship between the number of clusters and the WCSS (Within-Cluster Sum of Squares) value.

In this step, we calculate the WCSS values for various numbers of clusters, ranging from 1 to 10. The WCSS versus number of clusters graph is presented in Figure 2 below.

Figure 2. WCSS Number of Clusters Graph



In the graph, it can be seen that as the number of clusters increases, the WCSS value decreases significantly. However, this decline begins to slow down after reaching three clusters. The point where the WCSS value reduction starts to decrease significantly is called the elbow point. In this analysis, we found that the optimal number of clusters is 3.

From the analysis results, we were able to group students into three categories based on their assessment of Islamic religious knowledge. This clustering is expected to provide a clearer picture of students' understanding patterns in both aspects of knowledge, namely Fiqh (X1) and Aqidah Akhlak (X2).

With the determination of the optimal number of clusters, this research will proceed by applying the K-Means algorithm to group the students into these three clusters. Each cluster will represent groups of students with different levels of understanding of Islamic religious knowledge. The results of this clustering are expected to provide more useful information for educators in formulating more effective and responsive teaching strategies tailored to students' needs.

After applying the K-Means algorithm and grouping students into three clusters, the final step in this analysis is to evaluate the clustering results using the Silhouette Coefficient. The Silhouette Coefficient is a metric used to evaluate the quality of clustering by measuring how closely related objects within a cluster are compared to objects in other clusters.

The Silhouette Coefficient *(Si)* is calculated using two main parameters: *a(i)*, which is the average distance between data point i and all other points within the same cluster, and *b(i)*, which is the average distance between data point i and the nearest point in a different cluster. Based on the calculations, *a(i)* = 11.11 and *b(i)* = 24.56. Using the formula to calculate the Silhouette Coefficient, the resulting value is 0.548, which falls under the 'Good' category. This value indicates that the clustering performed is quite

effective, with students grouped within the same cluster showing good similarity in their Islamic religious knowledge, and the distance between clusters being relatively large.

With this good Silhouette Coefficient value, we can conclude that the K-Means algorithm successfully clustered the students accurately. This evaluation provides additional support for the previous analysis results and confirms that the separation of students into the 'High,' 'Medium,' and 'Low' clusters is representative and serves as a valid indicator for designing appropriate learning strategies.

Overall, this study demonstrates that the use of clustering techniques, particularly K-Means, can enhance the accuracy of assessing Islamic knowledge and provide deeper insights into students' educational needs. With these results, it is hoped that educators can formulate more effective and responsive learning approaches tailored to the individual needs of each student.

CONCLUSION

This study has successfully implemented the K-Means clustering technique to improve the accuracy of assessing Islamic knowledge among students at Madrasah in Kolaka, Indonesia. Based on the analysis conducted, several conclusions can be drawn as follows:

1. Data collected from 181 students indicate a significant variation in their understanding of two aspects of Islamic knowledge: Fiqh and Akidah Akhlak. The average scores show that although most students have a good understanding, there remains a group of students who require more attention.
2. Using the Elbow Method, this study successfully determined an optimal number of three clusters, representing different levels of understanding: "High," "Medium," and "Low." This cluster determination provides valuable insights for educators in assessing students' educational needs.
3. The application of the K-Means algorithm resulted in effective student clustering, with a balanced distribution among the three clusters. These findings indicate that clustering methods can be utilized to gain a deeper understanding of the learning dynamics and comprehension of students.
4. The Silhouette Coefficient value of 0.548, categorized as "Good," indicates that the clustering performed is valid and effective. This suggests that students within each cluster share a relatively good level of understanding.

Thus, this study concludes that the use of clustering techniques, particularly K-Means, can be an effective tool in educational assessment, providing a better understanding of student comprehension and assisting educators in developing more targeted and effective learning strategies. This research also opens avenues for further studies on the application of data analysis techniques in the educational context.

REFERENCES

Afifah, A. N., Nurcahyawati, V., & Hananto, V. R. (2023). Analisis Clustering dan Pemetaan Sebaran Pelanggan Perusahaan Properti di Sidoarjo. *Jurnal Edukasi Dan Penelitian Informatika*, *9*(3), 502–508. https://geokeo.com/database/state/id/

Badruttamam, A., Sudarno, S., & Maruddani, D. A. I. (2020). Penerapan Analisis Klaster K-Modes dengan Validasi Davies Bouldin Index Dalam Menentukan Klasterisasi Kanal Youtube di Indonesia (Studi Kasus: 250 Kanal YouTube Indonesia Teratas Menurut Socialblade). *Jurnal Gaussian*, *9*(3), 263–272. https://doi.org/10.14710/j.gauss.v9i3.28907

Das, D., Kayal, P., & Maiti, M. (2023). A K-means clustering model for analyzing the Bitcoin extreme value returns. *Decision Analytics Journal*, *6*(June 2022), 100152. https://doi.org/10.1016/j.dajour.2022.100152

Faizan, M., Zuhairi, M. F., Ismail, S., & Sultan, S. (2020). Applications of Clustering Techniques in Data Mining: A Comparative Study. *International Journal of Advanced Computer Science and Applications*, *11*(12), 146–153. https://doi.org/10.14569/IJACSA.2020.0111218

Fani, S. M., Santoso, R., & Suparti, S. (2021). Penerapan Text Mining Untuk Melakukan Clustering Data Tweet Akun Blibli Pada Media Sosial Twitter Menggunakan K-Means Clustering. *Jurnal Gaussian*, *10*(4), 583–593. https://doi.org/10.14710/j.gauss.v10i4.30409

Febianda, K., Ratnawati, D. E., & Rahayudi, B. (2020). Analisis Pengelompokkan Penjualan Rattan Furniture Pada PT. Hymsa Indotraco Berbasis Algoritma K-Means Clustering. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *4*(6), 1882–1887. http://j-ptiik.ub.ac.id

Guntara, M., & Lutfi, N. (2023). Optimasi Cacah Klaster pada Klasterisasi dengan Algoritma KMeans Menggunakan Silhouette Coeficient dan Elbow Method. *JuTI "Jurnal Teknologi Informasi," 2*(1), 43. https://doi.org/10.26798/juti.v2i1.944

Hartanti, N. T. (2020). Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *6*(2), 82–89. https://doi.org/10.25077/teknosi.v6i2.2020.82-89

Kurniadi, D., Agustin, Y. H., Akbar, H. I. N., & Farida, I. (2023). Penerapan Algoritma k-Means Clustering untuk Pengelompokan Pembangunan Jalan pada Dinas Pekerjaan Umum dan Penataan Ruang. *Aiti*, *20*(1), 64–77. https://doi.org/10.24246/aiti.v20i1.64-77

Maharrani, R. H., Abda'u, P. D., & Hastuti, D. H. (2022). JEPIN (Jurnal Edukasi dan Penelitian Informatika) Kabupaten Cilacap dengan Metode K-Means. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, *3*(3), 1–7. https://www.visualcrossing.com/.

Muningsih, E., Maryani, I., & Handayani, V. R. (2021). Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa. *Jurnal Sains Dan Manajemen*, *9*(1), 96. www.bps.go.id

Nabila, Z., Isnain, A. R., Permata, & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, *2*(2), 100. http://jim.teknokrat.ac.id/index.php/JTSI

Nugroho, B. S., Susilowati, Y., & Nurhasanah, N. (2022). Aplikasi Metode Analisis Fraktal dan K-Means Clustering untuk Identifikasi Retinopati Diabetik dan Retinopati Hipertensi menggunakan Citra Fundus Mata. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *8*(1), 128. https://doi.org/10.26418/jp.v8i1.54043

Orisa, M. (2022). Optimasi Cluster pada Algoritma K-Means. *Prosiding SENIATI*, *6*(2), 430–437. https://doi.org/10.36040/seniati.v6i2.5034

Oyelade, J., Isewon, I., Oladipupo, O., Emebo, O., Omogbadegun, Z., Aromolaran, O., Uwoghiren, E., Olaniyan, D., & Olawole, O. (2019). Data Clustering: Algorithms and Its Applications. *Proceedings - 2019 19th International Conference on Computational Science and Its Applications, ICCSA 2019*, *ii*, 71–81. https://doi.org/10.1109/ICCSA.2019.000-1

Paembonan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat. *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, *6*(2), 48. https://doi.org/10.51557/pt_jiit.v6i2.659

Pamuji, A. (2022). Data Mining dengan Segmentasi Pengguna pada Keamanan Sistem File. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *8*(1), 154. https://doi.org/10.26418/jp.v8i1.52233

Sopyan, Y., Lesmana, A. D., & Juliane, C. (2022). Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan. *Building of Informatics, Technology and Science (BITS)*, *4*(3), 1464–1470. https://doi.org/10.47065/bits.v4i3.2697

Sovia, R., Mandala, E. P. W., & Mardhiah, S. (2020). K-Means Algorithm in Selection of Outstanding Students and SAW Method for Prediction of Outstanding Scholarship Recipients. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *6*(2), 181. https://jurnal.untan.ac.id/index.php/jepin/article/view/37759

Srimurdianti, A., Sukamto, Setiawan, W., Esyudha, E., & Pratama. (2023). *JEPIN (Jurnal Edukasi dan Penelitian Informatika) Data Mining untuk Pengelompokan Saham pada Sektor Energi dengan Metode K-Means*. *9*(1), 76–81. https://jurnal.untan.ac.id/index.php/jepin/article/viewFile/62509/75676597024

Wahyu, A., & Rushendra, R. (2022). Klasterisasi Dampak Bencana Gempa Bumi Menggunakan Algoritma K-Means di Pulau Jawa. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *8*(1), 174. https://doi.org/10.26418/jp.v8i1.52260

Yang, M. S., & Hussain, I. (2023). Unsupervised Multi-View K-Means Clustering Algorithm. *IEEE Access*, *11*(January), 13574–13593. https://doi.org/10.1109/ACCESS.2023.3243133