

Mapping the Potential of Prospective New Students Utilizing K-Means and Fuzzy C-Means Clustering

Ahmad Fathoni^{1*}, Eri Zuliarso², Sarah Husain Toman³

^{1,2}Universitas Stikubank, Indonesia

³University of Al-Qadisiyah, Iraq

*e-mail: ahmadfathoni0001@mhs.unisbank.ac.id

ABSTRACT

UIN Walisongo Semarang has implemented outreach initiatives to engage prospective students, but these endeavors often lack data-informed approaches. This research intends to analyze the potential of incoming students through K-Means and Fuzzy C-Means (FCM) clustering techniques. The dataset comprises admission information from 2022 to 2024, focusing on five key factors: gender, admission pathway, study program, school type, and geographic origin. Data preprocessing was performed before conducting the clustering analysis. The Elbow Method and Silhouette Score were utilized to identify the optimal K for K-Means, whereas the Fuzzy Partition Coefficient and Xie-Beni Index were applied for FCM. Findings indicate that K-Means generated more distinct cluster boundaries, while FCM provided adaptability with overlapping clusters. Principal Component Analysis and the Davies-Bouldin Index were employed to facilitate the assessment. The mapping results are displayed by faculty, showcasing regional patterns and student demographics. This research establishes a data-driven basis for UIN Walisongo's strategic recruitment and admissions strategies.

Keywords:

Clustering; K-Mean; Fuzzy C-Means; Student Admission.

ABSTRAK

UIN Walisongo Semarang telah melaksanakan berbagai program sosialisasi untuk menarik calon mahasiswa baru, namun upaya ini sering kali belum berbasis data yang kuat. Penelitian ini bertujuan untuk memetakan potensi mahasiswa baru menggunakan algoritma klusterisasi K-Means dan Fuzzy C-Means (FCM). Dataset yang digunakan mencakup data penerimaan tahun

2022 hingga 2024, dengan fokus pada lima atribut utama: jenis kelamin, jalur penerimaan, program studi, jenis sekolah, dan daerah asal. Setelah dilakukan pra-proses data, analisis kluster dilakukan. Metode Elbow dan Skor Silhouette digunakan untuk menentukan nilai K optimal pada K -Means, sedangkan Fuzzy Partition Coefficient dan Indeks Xie-Beni digunakan untuk FCM. Hasilnya menunjukkan bahwa K -Means membentuk batas kluster yang lebih jelas, sementara FCM lebih fleksibel dengan tumpang tindih antar kluster. Evaluasi didukung oleh visualisasi Principal Component Analysis dan nilai Davies-Bouldin Index. Hasil pemetaan ditampilkan per fakultas, memberikan gambaran tren regional dan profil mahasiswa. Studi ini menjadi dasar strategis berbasis data dalam perencanaan promosi dan penerimaan mahasiswa baru.

Kata kunci:

Klusterisasi; K -Means; Fuzzy C-Means; Penerimaan Mahasiswa Baru.

1. Introduction

One of the main challenges faced by higher education institutions in Indonesia is the lack of alignment between student recruitment strategies and the actual characteristics of prospective applicants (Achmad & Widiastuti, 2024). While various outreach and promotional efforts are conducted each year, these initiatives are often based on generalized assumptions rather than empirical data. As a result, universities frequently encounter mismatches between promoted programs and the interests or backgrounds of incoming students. At UIN Walisongo Semarang, for instance, several study programs have experienced fluctuating enrollment patterns despite consistent promotional campaigns. This issue highlights the absence of a structured, data-informed approach to understanding who the prospective students are, where they come from, and what type of educational backgrounds they possess. Without such understanding, promotional efforts risk being ineffective and resource-intensive. Higher education is a crucial pillar in Indonesia's development, playing a key role in producing the next generation of quality individuals (Kusnadi & Ningrum, 2024). Universities are expected to be environments that help students cultivate their talents and interests through their chosen study programs (Liu, 2023). Therefore, effective strategies to promote higher education are essential to ensure that prospective students possess the appropriate potential and capabilities (Žalėnienė & Pereira, 2021). A well-designed outreach program can significantly improve student enrollment (Hairani, Susilowati, Puji Lestari, Marzuki, & Mardedi, 2022).

Several studies have demonstrated the application of clustering in education. For example, Han (2023) used Fuzzy C-Means (FCM) to cluster students based on levels of learning fatigue, revealing emotional and physical stress factors during continuous learning. Another study by Nabila et al. (2021) combined FCM with K-Nearest Neighbors (KNN) to predict on-time graduation among new students at UIN Sunan Ampel Surabaya. The results showed that factors such as how students were admitted, the type of school, their background in Islamic boarding schools, and their academic scores all affect

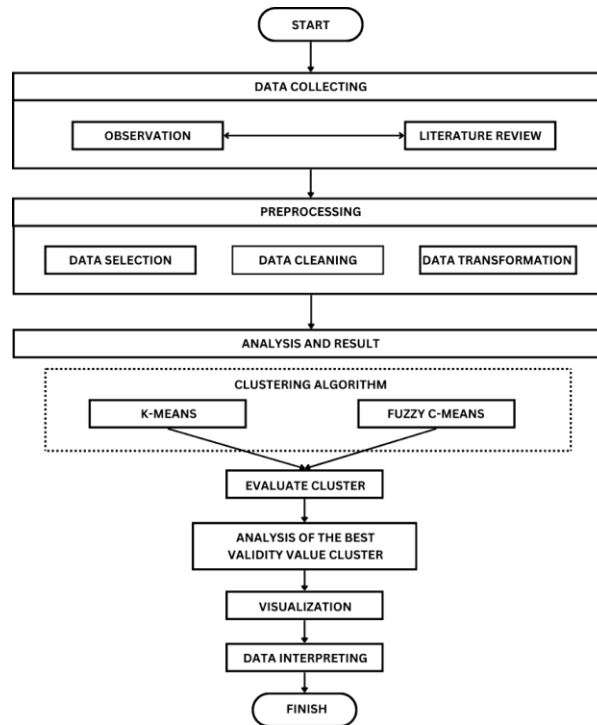
their graduation outcomes. Abdullah et al. (2022) utilized K-Means to categorize Indonesian provinces based on COVID-19 cases, recoveries, and deaths. Huang et al. (2020) combined K-Means with multiple linear regression to predict hospital admissions for chronic obstructive pulmonary disease (COPD), demonstrating improved accuracy and computational efficiency.

Despite the widespread use of clustering methods in educational data mining, there has been limited research specifically targeting the mapping of prospective new students before enrollment, particularly at UIN Walisongo Semarang. Although the university has conducted various outreach activities, it continues to face challenges in aligning program offerings with the actual characteristics and interests of applicants. The absence of a data-driven strategy has led to fluctuating enrollment patterns across study programs. This situation underscores a critical research problem: how can the university effectively map and understand the distribution and academic tendencies of its prospective students? Addressing this issue is essential for optimizing promotional strategies, resource allocation, and enrollment planning, especially in a geographically and socio-educationally diverse region like Central Java Province. This study uses data on prospective students from the 2022–2024 intake periods, obtained from the official admission records of UIN Walisongo Semarang. The dataset comprises 10,593 records, featuring key attributes such as gender, admission pathway, study program, school type, and region of origin.

2. Methods

This study employs a quantitative approach using unsupervised machine learning methods—specifically K-Means and Fuzzy C-Means (FCM)—to cluster and analyze the characteristics of prospective students at UIN Walisongo Semarang. Both algorithms have been widely used for pattern recognition and data segmentation in educational and non-educational domains (Ampadu, 2023; Rodriguez et al., 2019). The research process initiates with data collection, which comprises enrollment records spanning from 2022 to 2024. A total of 10,593 entries were gathered, all sourced from students in the Central Java Province. The chosen attributes encompass gender, admission pathway, accepted study program, type of previous institution, and the city or district of origin. Data preprocessing is performed to prepare the dataset for clustering. This process involves attribute selection, one-hot encoding for categorical variables, and normalization to ensure uniform scaling of the values. In the clustering phase, two algorithms are utilized. The K-Means algorithm, recognized for its straightforwardness and effectiveness, divides the data into clusters by reducing intra-cluster variance (Ikotun, Almutari, & Ezugwu, 2021). The ideal number of clusters is identified through the Elbow Method and the Silhouette Score (Rodriguez et al., 2019). The second method, Fuzzy C-Means (FCM), allows for degrees of membership in multiple clusters (Zhou, Li, Zhang, & Ping, 2021), making it suitable for ambiguous data (Han, 2023). The performance of FCM is evaluated using the Fuzzy Partition Coefficient (FPC) and the Xie-Beni Index, which assess the quality and compactness of clusters (Al-Alawi, Jaddoa, Cugley, & Hassanin, 2024). Next, an evaluation phase is performed. Principal Component Analysis (PCA) is used to reduce dimensionality and visualize cluster patterns more clearly (Huang et al., 2020). Furthermore, the Davies-Bouldin Index (DBI) is employed to validate the quality of clusters based on intra- and inter-cluster distances (Abdullah et al., 2022).

Finally, the resulting clusters are visualized in the result mapping stage, showing the distribution of prospective students across different faculties and regions. These results are expected to support data-driven decisions in the university's promotional and admission strategies. The methodology flow is summarized in a flowchart (see Picture 1).



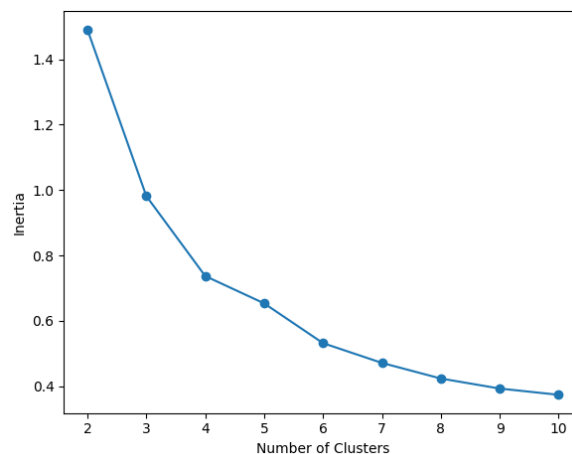
Picture 1. Research Method Flowchart

2.1 Data Processing

2.1.1 K-Means Clustering

K-Means is a prominently utilized hard clustering method (Ran, Zhou, Lei, Tepsan, & Deng, 2021) recognized for its simplicity and efficacy in handling extensive datasets (Singh & Singh, 2024; Zaib & Ourabah, 2023). Its straightforward algorithm and scalability contribute to its widespread application in various fields of research and data analysis (Patel & Kushwaha, 2020). This method clusters the data into K groups by minimizing variation within each cluster (Chong, 2021). The center of each cluster is determined as the mean of all data points within that cluster. This process is iterated until there is no significant change in the cluster centers. The first steps in applying the K-Means method involve determining the number of clusters (K) and then randomly selecting K initial centroid points. Then, calculate the distance of each data point to these centroids and assign each data point to the cluster that is nearest to it. After assigning the data points, update the centroid positions by recalculating the mean of all data points in each cluster. Repeat this process of calculating distances and reassigning data points until the algorithm converges, meaning there is no significant change in the positions of the centroids (Ahmed, Seraj, & Islam, 2020; Chong, 2021; Miraftebzadeh, Colombo, Longo, & Foidelli, 2023). The researcher employed two systematic methods to determine the optimal number of K clusters: the Elbow Method and the Silhouette Score. The Elbow Method is a

widely recognized technique used to ascertain the ideal number of clusters in the K-Means Clustering algorithm (Sadeghi, Casey, Carranza, & Lynch, 2024), thereby enhancing the decision-making process in cluster analysis (Ling & Weiling, 2025). The Elbow Method is a statistical approach utilized to ascertain the optimal number of clusters (K) for a given dataset in cluster analysis (Pampuch, Negri, Loikith, & Bortolozo, 2023). This method involves calculating the Within-Cluster Sum of Squares (WCSS) (Huang et al., 2020; Motlagh, Berry, & O'Neil, 2019), which quantifies the total squared Euclidean distance between each data point and its corresponding cluster centroid across varying values of K. As K increases, the WCSS is expected to decrease (Yin, Yong, Qi, Yang, & Lan, 2023), reflecting the phenomenon that a greater number of clusters results in reduced distances between data points and their centroids (Chin & Collins, 2024). The objective is to identify the "elbow point" in the WCSS plot, where the rate of decrease in WCSS sharply changes, indicating a suitable balance between model complexity and variance (Sammouda & El-Zaart, 2021). In the Elbow Method, the Within-Cluster Sum of Squares (WCSS) is calculated for a range of cluster numbers (K) from 1 to 10. The WCSS values are plotted against the corresponding K values, producing a graphical representation. This graph typically illustrates a downward curve, indicating that as the number of clusters (K) increases, the WCSS generally decreases. The point at which the rate of decrease significantly slows, forming an "elbow," is often considered the optimal number of clusters for the given dataset (Et-Taleby, Boussetta, & Benslimane, 2020). The point in a chart where the rate of change begins to diminish, resulting in a distinct angle or "elbow," is referred to as the elbow point (Shi et al., 2021). This point indicates the ideal number of clusters (Shi et al., 2021).



Picture 2. Elbow Method K-Means Clustering

Picture 2. illustrates the Elbow Method graph, which displays a significant decrease in inertia as the number of clusters (k) increases. Inertia, which measures the variance within a cluster, exhibits a sharp decline from k = 2 to k = 4. After k = 4, the decrease in inertia becomes more gradual and continues to level off towards k = 10. The calculations suggest that the optimal number of clusters is either k = 3 or k = 4. Further calculations using the Silhouette Score need to be conducted to confirm this (Habsy, Zakirah, Rahmah, & Nafisah, 2023). The results of the Elbow Method and Silhouette Score calculations are shown in Picture 3.

	Number of Clusters	Elbow Inertia	Silhouette Score
0	2	1.491096e+06	0.335421
1	3	9.817664e+05	0.368241
2	4	7.371241e+05	0.355693
3	5	6.535650e+05	0.336687
4	6	5.318495e+05	0.318734
5	7	4.714126e+05	0.320422
6	8	4.236190e+05	0.303323
7	9	3.929481e+05	0.289746
8	10	3.737915e+05	0.275509

Optimal number of clusters (Silhouette): 3

Picture 3. Silhouette Score K-Means Clustering

The highest silhouette value was observed at $k = 3$, with a score of 0.368241. After $k=3$, the score sharply declines at $k=4$ and continues to fluctuate before gradually declining further as the number of clusters increases. Although the Elbow method identifies $k=4$ as a potential cluster number, the relatively lower silhouette value for $k=4$ compared to $k=3$ indicates that $k=3$ may be a more optimal choice in terms of clustering quality based on both metrics. Considering these two methods, the researcher selected $k = 3$ for K-Means clustering. Table 1 presents the results of the K-Means clustering data processing.

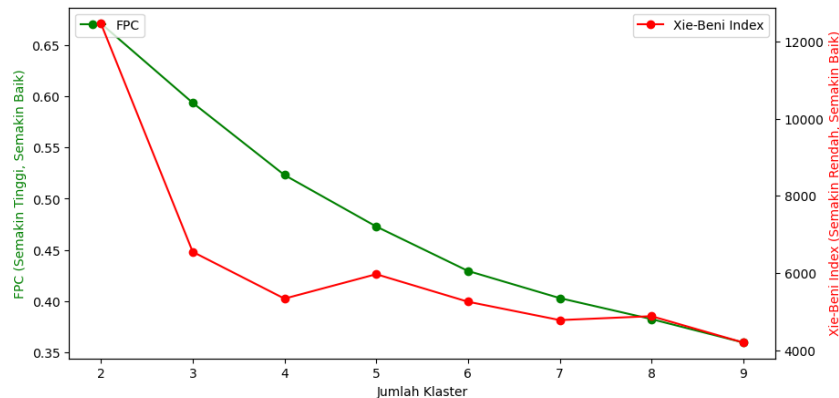
Table 1. Number of Members in Each K-Means Cluster

cluster	Number of
0	3.945
1	3.015
2	3.633
Total	10.593

2.1.2 FCM Clustering

Fuzzy C-Means (FCM) is a partitioning clustering algorithm (Rajkumar, Yesubabu, & Subrahmanyam, 2019) That employs a data grouping technique in which a membership value represents the degree of membership of each data point to a particular cluster (Lei et al., 2018). This value quantifies the extent to which a data point belongs to each cluster, allowing for the formation of fuzzy clusters where an individual data point can simultaneously belong to multiple clusters with varying degrees of certainty (Rayala & Kalli, 2021). The main goal of Fuzzy C-Means (FCM) is to partition objects in multiple dimensions into a specified number of clusters (Hassan, Shah, Othman, & Hassan, 2020). This process aims to find centroids that minimize the disparity between the clusters (Dubey, Gupta, & Jain, 2018). FCM is classified as a soft clustering method (Gatea & Al-asadi, 2024; Krasnov et al., 2023), which means that objects can belong to more than one cluster with varying degrees of membership. Objects positioned at the boundaries of clusters are not exclusively assigned to a single cluster; instead, they may exhibit partial membership across multiple clusters (Jia & Cheung, 2018), characterized by membership degrees ranging from 0 to 1. The selection of the optimal number of clusters (K) is a critical parameter in the clustering process (Qin et al., 2022), as it directly impacts the efficacy of the segmentation results in accurately representing the intrinsic

structure of the data. This study employed two quantitative methods to evaluate cluster validity: the Fuzzy Partition Coefficient (FPC) and the Xie-Beni Index (XB). These methodologies provide a systematic framework for assessing the appropriateness and robustness of the identified clusters (Zhang, Wu, Xu, Ning, & Han, 2025).



Picture 4. Fuzzy Partition Coefficient (FCP) and Xie-Beni Index Graph

Picture 4 illustrates the relationship between the number of clusters and the values of the Fuzzy Partition Coefficient (FPC) and the Xie-Beni (XB) index. The graph shows that as the number of clusters increases, the FPC value consistently decreases, while the XB values fluctuate but generally trend downward. The FPC (indicated in green) drops dramatically after $K=2$, remains relatively stable at $K=3$ and $K=4$, then continues to decline. Conversely, the XB (shown in red) experiences a sharp decline from $K=2$ to $K=3$, followed by stabilization from $K=7$ to $K=9$. The ideal candidate, based on a balance of high FPC and low XB, is either $K=3$ or $K=4$.

Table 2. Comparison of FCP and XB Values

Number of Clusters	Fuzzy Partition Coefficient (FPC)	Xie-Beni Index (XB)
2	0.670976	12460.074195
3	0.593590	6551.846392
4	0.523101	5341.295929
5	0.472851	5973.326969
6	0.429467	5260.983535
7	0.402967	4785.785023
8	0.382462	4886.892488
9	0.359633	4205.976060

Table 2 indicates that the highest FPC value was achieved at $K=2$ (0.670976), while the lowest XB value was recorded at $K=9$ (4205.976060). To identify the optimal number of clusters, it is essential to strike a balance between these two metrics. $K=3$ or $K=4$ are better candidates because they both show a relatively high FPC value and an XB value that is beginning to stabilize. The

researcher chose $K = 3$ for the FCM analysis due to its balanced FPC and XB values. Table 3 presents the results of the FCM clustering.

Table 3. Number of FCM Cluster Members

Cluster	Number of members
0	3.913
1	3.587
2	3.093
Total	10.593

3. Results and Discussion

The results of K-Means and FCM clustering were analyzed to identify the characteristics of each cluster based on five indicators.

3.1 Main Characteristics of K-Means Clustering

Cluster 0 in the K-Means clustering results exhibits the strongest connection to Islamic education, highlighted by the significant number of female students. Most students in this cluster come from Islamic High School (MA), indicating a prior religiously oriented educational background. In terms of study programs, many students in this cluster opt for disciplines such as Islamic Religious Education (PAI), Elementary Madrasah Teacher Education (PGMI), Psychology (PSI), and Islamic Banking (PBS). This suggests a tendency towards careers in Islamic-based education and social services. The primary entry route for these students is through UM-PTKIN, followed by SPAN-PTKIN and the independent admission route (Mandiri), indicating that the majority are admitted via the national academic selection for PTKIN. The highest number of students in this cluster originates from Kendal, Demak, Pati, Jepara, and Grobogan Regencies, regions known for their strong presence of Madrasah Aliyah in Central Java Province. Cluster 1 presents a more balanced distribution of students from varied educational backgrounds. The origins of the schools are relatively diverse, making this cluster more heterogeneous compared to others. In terms of study programs, students are not concentrated in a single discipline; instead, they are spread across various fields, including Management (MAN), Islamic Communication and Broadcasting (KPI), Law (IH), and Islamic Religious Education (PAI), with no particular program overwhelmingly prominent. This indicates that this cluster comprises students with a broad range of interests, particularly in social sciences, management, and Islamic law. The distribution of students by districts and cities shows a predominance of urban areas. Many students in this cluster hail from Semarang City, Tegal City, and Tegal Regency, as well as Rembang and Temanggung Regencies. Regarding the entrance routes, students are evenly distributed across UM-PTKIN, Mandiri, and SPAN-PTKIN, demonstrating flexibility in their admission processes. Cluster 2 exhibits a strong focus on the fields of economics and Islamic law. The number of female students surpasses that of male students, with the majority having graduated from Madrasah Aliyah; however, there are also graduates from high schools and vocational schools. The most popular study programs include Sharia Accounting, Sharia Economics, Law, and Islamic Criminal Law, highlighting an academic emphasis on finance and the application

of Islamic law. This cluster appears to be more technocratic and practical in orientation compared to others. Many of the students in this cluster come from areas such as Kendal, Demak, Pati, Pemasang, and Brebes Regencies, which are not only known for their Islamic educational institutions but also feature developing community economic activities. The pathways for student entrance in this cluster are also evenly divided among UM-PTKIN, Mandiri, and SPAN-PTKIN, indicating that students come from diverse and competitive backgrounds.

Table 4. Characteristics of Each K-Means Cluster Based on Five Indicators

Indicator	Cluster 0	Cluster 1	Cluster 2
Gender	Dominated by females (2,843); males (1,102)	female (1.886); male (1.129)	female (2.266); male (1.367)
Study Program (big 5)	Islamic Religious Education (PAI), Elementary Madrasah Teacher Education (PGMI), Psychology (PSI), Islamic Banking (PBS), Sosiologi	Islamic Religious Education (PAI), Management (MAN), Islamic Communication and Broadcasting (KPI), Law (IH)	Sharia Accounting (AKS), Islamic Economy (EI), Law (IH), Islamic Criminal Law (HPI)
Type of School (big 5)	MA (2.433), lalu SMA (1.098), SMK (367)	MA (1.493), SMA (1.009), SMK (447)	MA (2.032), SMA (1.113), SMK (443)
Regency/City (big 5)	Kendal, Demak, Pati, Jepara, Grobogan	Semarang City, Tegal, Rembang, Temanggung	Kendal, Demak, Pati, Pemasang, Brebes
Admission Pathway (3 besar)	UM-PTKIN, SPAN-PTKIN, Mandiri	UM-PTKIN, Mandiri, SPAN-PTKIN	UM-PTKIN, Mandiri, SPAN-PTKIN

3.2 Main Characteristics of FCM Clustering

Cluster 0 primarily consists of female students, with most of them coming from Madrasah Aliyah (MA), indicating a strong connection to Islamic-based education. Students in this cluster tend to choose study programs such as Islamic Religious Education, Madrasah Ibtidaiyah Teacher Education (PGMI), and Psychology, which reflect their focus on educational and social fields. The majority of students gain admission through the University of Michigan-PTKIN. Most of the students in this cluster are from Kendal, Demak, and Pati Regencies, regions known for a high number of madrasah graduates. This suggests that Cluster 0 is dominated by students with a robust Islamic education background, who are inclined to pursue fields in education and social studies, and typically enter through merit-based selection and PTKIN national exams. In contrast, Cluster 1 shows a more balanced gender composition. This cluster's study programs are more oriented toward Islamic economics, law, and management. While most students still come from Madrasah Aliyah (MA), the proportion of high school graduates is higher in this cluster compared to Cluster 0. In terms of regional origin, many students in this cluster come from the Kendal, Demak, Pati, and Grobogan Regencies. The admission paths for this cluster show a more diverse acceptance pattern, with students entering through UM-PTKIN, Independent Selection, and SPAN-PTKIN. This diversity suggests that students in this cluster have a stronger interest in the fields of economics and Islamic law, as well as a more varied educational background, compared to other clusters. Cluster 2 comprises both female and male students, with a more balanced gender ratio than Cluster 0. In terms of study programs, students from this cluster predominantly choose Sharia Accounting, Sharia Economics, Legal Science, and Islamic

Criminal Law, reflecting a greater interest in Islamic law and finance. The educational backgrounds of these students vary, with the majority graduating from Madrasah Aliyah, high schools, and vocational schools. Regionally, many students are from Kendal, Demak, Pati, Pemalang, and Brebes Regencies, which have large populations and relatively good access to education. The admission pathways for this cluster are similar to those of Cluster 1, with most students accepted through UM-PTKIN, Independent Selection, and SPAN-PTKIN. Overall, this cluster is characterized by students who are highly focused on economics and Islamic law, possess diverse educational backgrounds, and exhibit a balanced approach to acceptance, combining exam-based selection processes with academic achievements.

Table 5. Characteristics of Each FCM Cluster Based on Five Indicators

Indicator	Cluster 0	Cluster 1	Cluster 2
Gender	Dominated by females (2.821), males (1.092)	female (2.238); male (1.349)	female (2.266); male (1.367)
Study Program (big 5)	Islamic Religious Education (PAI), Elementary Madrasah Teacher Education (PGMI), Psychology (PSI), Islamic Banking (PBS), Sosiologi	Akuntansi Syariah, Ekonomi Syariah, Manajemen, Hukum Keluarga Islam	Akuntansi Syariah, Ekonomi Syariah, Ilmu Hukum, Hukum Pidana Islam
Type of School (big 5)	MA (2.413), lalu SMA (1.090), SMK (363)	MA (2.015), SMA (1.093), SMK (436)	MA (2.032), SMA (1.113), SMK (443)
Regency/City (big 5)	Kendal, Demak, Pati, Jepara, Grobogan	Kendal, Demak, Pati, Grobogan, Jepara	Kendal, Demak, Pati, Pemalang, Brebes
Admission Pathway (big 3)	UM-PTKIN (1.683), SPAN-PTKIN (748), Mandiri (590)	UM-PTKIN (1.379), Mandiri (668), SPAN-PTKIN (556)	UM-PTKIN (1.397), Mandiri (677), SPAN-PTKIN (560)

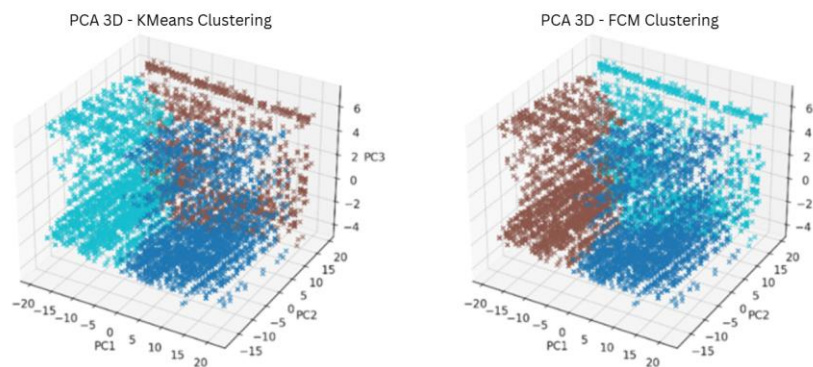
3.3 Evaluation and Comparison of K-Means and FCM

In clustering analysis, the process of dividing data into distinct groups often presents various challenges that can impact the effectiveness of the resulting clusters. It is essential to evaluate several key aspects to ensure meaningful separation: (1) quality of cluster separation, are the clusters formed with clear, well-defined boundaries, or is there significant overlap where one cluster encroaches upon another? The clarity of these boundaries is crucial for accurately distinguishing between different groups. (2) compactness of clusters; how closely packed is the data within each cluster? Is the data homogeneous, with similar characteristics, or is it scattered across a broader range of values? A compact cluster indicates that the items within it share more similarities, improving the cluster's reliability. (3) relationships between clusters; what is the distance between the different clusters? Are they sufficiently separated so that each segment stands alone, or are they too close together, risking the potential for confusion? Proper separation is necessary to ensure that each cluster represents a distinct category. To address these challenges, this evaluation employs two primary approaches, one

of which involves dimensionality reduction through Principal Component Analysis (PCA). This technique helps simplify the data's structure while retaining essential information, making it easier to analyze and interpret clustering results (Gewers et al., 2021) and clustering quality measurement using the Davies-Bouldin Index (DBI) (Mughnyanti, Efendi, & Zarlis, 2020).

3.3.1 Principal Component Analysis (PCA)

The results of the K-Means and FCM clustering were visually represented through a dimensionality reduction process using Principal Component Analysis (PCA). This method projects high-dimensional data, based on five indicators: gender, admission path, study program, district or city of origin, and type of school, into a three-dimensional space (PC1, PC2, and PC3). The results indicate that the clusters formed by K-Means and FCM can be distinguished more clearly in this 3D space, despite some overlapping structures.



Picture 5. 3D Principal Component Analysis (PCA) for K-Means and Fuzzy C-Means (FCM)

Based on the visualization results, several key differences are observed between K-Means and FCM clustering in terms of data distribution and cluster segmentation. K-Means clustering demonstrates a more definitive separation between data groups. Each cluster is distinctly segmented with clear boundaries in three-dimensional space (PC1, PC2, PC3). This observation confirms that K-Means utilizes hard clustering, meaning each data point can belong exclusively to one cluster. In contrast, the results of FCM clustering exhibit greater flexibility, with smoother transitions between clusters. In the 3D PCA visualization, the data points exhibit interpolation between clusters, a fundamental characteristic of the fuzzy clustering approach.

3.3.2 Davies Bouldin Index (DBI)

In addition to performing a visual assessment using Principal Component Analysis (PCA), the quality of clustering was quantitatively evaluated through the Davies-Bouldin Index (DBI). The DBI quantifies the ratio of intra-cluster compactness to inter-cluster separation, serving as an indicator of clustering performance. Table 6 presents a comparative analysis of the DBI values derived from the K-Means and FCM clustering algorithms. A lower DBI value signifies superior clustering quality, as it reflects a greater distinction between clusters and enhanced separation relative to the internal dispersion of each cluster.

Table 6. DBI Values Obtained from K-Means and FCM Clustering

Clustering Method	Davies-Bouldin Index (DBI)
K-Means	0.9269
FCM	0.9322

The application of the K-Means clustering algorithm yielded a Davies-Bouldin Index (DBI) of 0.9269, suggesting that it facilitates the formation of clusters that are both compact and well-separated. In the K-Means framework, each data point is assigned to a single cluster (a method termed hard clustering), which enhances the clarity of the boundaries between clusters. This configuration results in reduced average intra-cluster distances and increased inter-cluster distances compared to the FCM algorithm. Consequently, the separation between clusters is more pronounced, diminishing the probability of data points residing in transitional areas between clusters. This observation is corroborated by the 3D PCA visualization, which demonstrates that the K-Means clustering results exhibit more distinct and delineated cluster structures. In contrast, the FCM algorithm produced a Davies-Bouldin Index (DBI) of 0.9322, which is slightly higher than the value obtained from the K-Means algorithm. This indicates that clustering with FCM results in clusters characterized by enhanced flexibility and increased overlap. The FCM methodology enables each data point to possess a degree of membership across multiple clusters, allowing for the presence of certain data points within transitional zones. While this flexibility permits FCM to capture more intricate data patterns effectively, it simultaneously results in a less distinct separation between clusters when compared to K-Means. The three-dimensional Principal Component Analysis (PCA) visualization of the FCM clusters further illustrates this phenomenon, revealing considerable overlap, particularly in the transitional regions between divergent cluster groups.

3.4 Mapping of Potential Students

This study identified three main clusters among prospective new students using both K-Means and Fuzzy C-Means algorithms. While both methods yielded consistent patterns, K-Means provided more distinct cluster boundaries, whereas FCM allowed for a degree of overlap in student characteristics. The discussion below focuses on two contrasting faculties—the Faculty of Science and Technology (FST) and the Faculty of Ushuluddin and Humanities (FUHUM)—as representative case studies based on their differing student profiles and educational backgrounds.

Table 7. Mapping of The Faculty of Science and Technology Students

Study Program	Regency/City of Origin (number)	K-Means Cluster	FCM Cluster
Biology Education (PB)	Kab. Pati (25), Kab. Demak (24), Kab. Jepara (23), Kab. Kendal (22), Kab. Pemalang (21), Kota Semarang (19), Kab. Brebes (18), Kab. Grobogan (16), Kab. Blora (13), Kab. Tegal (13), Kab. Batang (12), Kab. Kudus (12), Kab. Temanggung (6), Kab. Rembang (6), Kab. Kebumen (6), Kab. Banjarnegara (6), Kab. Boyolali (5), Kab. Cilacap	0, 1	0, 2

	(5), Kab. Klaten (4), Kab. Semarang (4), Kota Tegal (4), Kab. Wonosobo (3), Kab. Magelang (3), Kab. Sragen (3), Kab. Pekalongan (3), Kab. Purworejo (3), Kota Pekalongan (2), Kab. Wonogiri (2), Kab. Sukoharjo (2), Kab. Purbalingga (2), Kota Magelang (2), Kab. Banyumas (1)		
Physics Education (PF)	Kab. Pati (7), Kab. Jepara (7), Kab. Kendal (6), Kab. Tegal (4), Kab. Kudus (4), Kab. Demak (3), Kab. Semarang (3), Kab. Grobogan (3), Kab. Pemalang (3), Kota Semarang (2), Kab. Wonosobo (2), Kab. Batang (2), Kab. Sragen (2), Kab. Boyolali (1), Kab. Cilacap (1), Kab. Banyumas (1), Kab. Kebumen (1), Kab. Purbalingga (1), Kab. Banjarnegara (1), Kab. Rembang (1), Kota Tegal (1)	0, 1	0, 2
Chemical Education (PK)	Kab. Pati (15), Kab. Demak (13), Kab. Grobogan (10), Kab. Brebes (8), Kab. Jepara (7), Kab. Kudus (6), Kota Semarang (6), Kab. Kendal (6), Kab. Tegal (4), Kab. Cilacap (4), Kab. Boyolali (4), Kab. Klaten (3), Kab. Pekalongan (3), Kab. Banyumas (3), Kab. Pemalang (3), Kab. Semarang (3), Kab. Temanggung (2), Kab. Rembang (2), Kota Tegal (2), Kab. Banjarnegara (2), Kab. Kebumen (1), Kab. Magelang (1), Kab. Purbalingga (1), Kab. Sragen (1), Kab. Blora (1), Kab. Wonosobo (1), Kab. Batang (1)	0, 1	0, 2
Mathematics Education (PM)	Kab. Pati (27), Kab. Jepara (22), Kab. Kendal (21), Kota Semarang (19), Kab. Demak (19), Kab. Rembang (16), Kab. Tegal (15), Kab. Grobogan (12), Kab. Blora (12), Kab. Kudus (11), Kab. Brebes (10), Kab. Temanggung (9), Kab. Semarang (6), Kab. Kebumen (5), Kab. Batang (5), Kab. Cilacap (5), Kab. Pekalongan (4), Kab. Wonogiri (3), Kab. Sragen (3), Kab. Wonosobo (3), Kab. Banjarnegara (3), Kab. Pemalang (3), Kota Tegal (2), Kota Pekalongan (2), Kota Salatiga (2), Kab. Magelang (2), Kab. Banyumas (2), Kab. Boyolali (2), Kota Surakarta (1), Kab. Purworejo (1), Kab. Karanganyar (1), Kab. Klaten (1), Kab. Sukoharjo (1)	1, 2	1, 2
Biology (BIO)	Kota Semarang (23), Kab. Kendal (17), Kab. Demak (15), Kab. Brebes (13), Kab. Pemalang (12), Kab. Pati (12), Kab. Grobogan (11), Kab. Tegal (9), Kab. Rembang (7), Kab. Kudus (7), Kab. Batang (7), Kab. Blora (6), Kab. Semarang (3), Kab. Purbalingga (3), Kab. Jepara (3), Kab. Magelang (3), Kab. Wonosobo (3), Kab. Sragen (2), Kab. Cilacap (2), Kab. Karanganyar (2), Kab. Klaten (2), Kab. Kebumen (2), Kota Tegal (2), Kab. Purworejo (2), Kab. Temanggung (1), Kab. Banjarnegara (1), Kab. Boyolali (1), Kab. Banyumas (1), Kab. Pekalongan (1), Kab. Wonogiri (1)	1, 2	1, 2
Physics (FIS)	Kab. Kendal (5), Kab. Pati (5), Kota Semarang (5), Kab. Demak (4), Kab. Jepara (3), Kab. Grobogan (3), Kab. Kudus (3), Kab. Brebes (3), Kab. Rembang (2), Kab. Purbalingga (2), Kab. Semarang (2), Kab. Pemalang (1), Kab. Sragen (1), Kab. Klaten (1), Kab. Purworejo (1), Kota Pekalongan (1), Kab. Tegal (1), Kab. Cilacap (1), Kota Tegal (1), Kab. Wonosobo (1), Kab. Batang (1)	1, 2	1, 2
Chemistry (KIM)	Kota Semarang (15), Kab. Kendal (13), Kab. Kudus (13), Kab. Jepara (11), Kab. Demak (10), Kab. Pati (9), Kab. Brebes (6), Kab. Grobogan (5), Kab. Tegal (5), Kab. Pemalang (4), Kab. Semarang (4), Kab. Batang (4), Kab. Boyolali (3), Kab. Pekalongan (2), Kab. Banjarnegara (2), Kab. Kebumen (2), Kab. Klaten (1), Kab. Temanggung (1), Kab. Magelang (1), Kab. Wonosobo (1), Kab. Wonogiri (1), Kab. Karanganyar (1), Kab. Rembang (1), Kab. Blora (1)	0, 1	0, 2

Mathematics (MAT)	Kab. Kendal (25), Kota Semarang (19), Kab. Jepara (16), Kab. Demak (10), Kab. Grobogan (8), Kab. Rembang (8), Kab. Pati (7), Kab. Pemalang (7), Kab. Semarang (6), Kab. Brebes (6), Kab. Batang (5), Kab. Boyolali (4), Kab. Cilacap (4), Kab. Tegal (4), Kab. Kudus (3), Kab. Pekalongan (2), Kab. Temanggung (2), Kota Tegal (2), Kab. Wonosobo (2), Kab. Kebumen (2), Kab. Blora (2), Kab. Banyumas (2), Kab. Wonogiri (1), Kab. Magelang (1), Kab. Banjarnegara (1), Kab. Purbalingga (1), Kota Pekalongan (1)	0, 1	0, 2
Environmental Engineering (TL)	Kota Semarang (41), Kab. Kendal (17), Kab. Demak (11), Kab. Grobogan (10), Kab. Pati (9), Kab. Tegal (7), Kab. Brebes (7), Kab. Pemalang (5), Kab. Jepara (4), Kab. Pekalongan (3), Kab. Semarang (3), Kab. Kudus (3), Kab. Batang (3), Kab. Blora (3), Kab. Rembang (3), Kab. Temanggung (2), Kab. Kebumen (2), Kab. Sukoharjo (1), Kab. Boyolali (1), Kab. Sragen (1), Kab. Wonogiri (1), Kota Pekalongan (1), Kota Tegal (1), Kab. Karanganyar (1), Kab. Purworejo (1), Kab. Cilacap (1)	0, 1	0, 2
Information Technology (TI)	Kota Semarang (65), Kab. Kendal (29), Kab. Jepara (26), Kab. Demak (20), Kab. Pati (18), Kab. Pemalang (18), Kab. Blora (13), Kab. Kudus (12), Kab. Grobogan (10), Kab. Brebes (9), Kab. Batang (7), Kota Pekalongan (6), Kab. Tegal (6), Kab. Cilacap (6), Kab. Rembang (5), Kab. Wonosobo (5), Kab. Boyolali (4), Kab. Temanggung (4), Kab. Magelang (4), Kab. Wonogiri (4), Kab. Banyumas (4), Kab. Purbalingga (3), Kab. Kebumen (3), Kab. Klaten (3), Kab. Banjarnegara (3), Kab. Pekalongan (3), Kab. Semarang (3), Kota Surakarta (2), Kota Tegal (2), Kab. Sukoharjo (2), Kab. Karanganyar (1), Kab. Sragen (1)	0, 1	0, 2

The clustering analysis revealed that FST primarily attracts students from general senior high schools (SMA) rather than Islamic educational institutions. This is consistent with the type of programs available, which are generally STEM-focused and tend to attract students with secular academic backgrounds. In Cluster 1, for instance, more than 70% of students admitted to FST programs originated from general schools, with a notable portion gaining entry through the achievement-based admission pathway. This confirms findings by Hairani et al. (2022), who noted that students' academic orientation significantly influences the program's attractiveness. Moreover, this trend resonates with the analysis by Han, (2023), which found that admission path and school type were crucial predictors in modeling academic success. The data indicated that FST applicants often come from urban areas with robust public high school systems, highlighting the significance of geographic focus in promotional strategies (Marbouti, Ulas, & Wang, 2021).

In comparison, the future students of FUHUM primarily come from Islamic educational institutions (MA) and pesantren-based settings, with a significant number gaining entry through regular or exam-based admission routes. Cluster 2, reflecting a substantial portion of applicants heading to FUHUM, indicated robust backgrounds in religious education and was largely composed of female students. This pattern supports (Sahin, 2018) The Assertion that students' interests in the humanities and religious studies are often shaped by their earlier educational environments. The dominance of female applicants also echoes previous observations about gendered preferences in

program selection (Zarkasyi, 2024). In terms of regional distribution, FUHUM receives more students from semi-rural areas, especially those with strong Islamic educational traditions. These insights underline the need for faith-based promotion strategies that resonate with the values and experiences of potential applicants.

Table 8. Mapping of the Faculty of Ushuluddin dan Humanities Students

Study Program	Regency/City of Origin (number)	K-Means Cluster	FCM Cluster
Islamic Creed and Philosophy (AFI)	Kota Semarang (30), Kab. Demak (21), Kab. Kendal (19), Kab. Pati (17), Kab. Tegal (11), Kab. Grobogan (9), Kab. Brebes (9), Kab. Batang (8), Kab. Kebumen (5), Kab. Pemalang (5), Kab. Kudus (5), Kab. Semarang (5), Kota Tegal (4), Kab. Jepara (4), Kab. Rembang (4), Kab. Boyolali (2), Kab. Purbalingga (2), Kota Pekalongan (2), Kab. Magelang (1), Kab. Blora (1), Kab. Cilacap (1), Kab. Banjarnegara (1), Kab. Temanggung (1), Kab. Purworejo (1), Kab. Wonosobo (1), Kota Magelang (1), Kab. Pekalongan (1)	1, 2	1, 2
Qur'an Science and Interpretation (IAT)	Kab. Demak (45), Kota Semarang (40), Kab. Kendal (33), Kab. Tegal (24), Kab. Jepara (22), Kab. Pati (21), Kab. Rembang (17), Kab. Pemalang (16), Kab. Semarang (13), Kab. Brebes (12), Kab. Grobogan (11), Kab. Batang (11), Kab. Kudus (10), Kab. Blora (8), Kab. Banjarnegara (6), Kab. Purbalingga (6), Kota Pekalongan (6), Kab. Kebumen (5), Kota Tegal (4), Kab. Pekalongan (4), Kab. Banyumas (3), Kab. Klaten (3), Kab. Magelang (3), Kab. Wonosobo (3), Kab. Temanggung (2), Kab. Wonogiri (2), Kab. Cilacap (2), Kota Magelang (1), Kab. Purworejo (1), Kab. Sragen (1), Kab. Sukoharjo (1)	1, 2	1, 2
Islamic Arts and Architecture (ISAI)	Kota Semarang (25), Kab. Demak (21), Kab. Tegal (18), Kab. Kendal (17), Kab. Pati (15), Kab. Jepara (15), Kab. Grobogan (13), Kab. Pekalongan (8), Kab. Semarang (7), Kab. Kudus (7), Kab. Brebes (7), Kab. Pemalang (6), Kab. Temanggung (5), Kab. Rembang (5), Kab. Blora (4), Kab. Sragen (4), Kab. Karanganyar (3), Kab. Magelang (3), Kab. Boyolali (3), Kab. Batang (3), Kab. Banyumas (3), Kab. Banjarnegara (2), Kota Pekalongan (2), Kab. Kebumen (2), Kota Tegal (2), Kab. Klaten (1), Kab. Wonogiri (1), Kab. Wonosobo (1), Kab. Purbalingga (1)	1, 2	1, 2
Religious Studies	Kota Semarang (10), Kab. Kendal (9), Kab. Pemalang (5), Kab. Rembang (5), Kab. Jepara (4), Kab. Grobogan (4), Kab. Demak (4), Kab. Semarang (3), Kab. Brebes (3), Kab. Tegal	0, 1	0, 2

(SAA)	(3), Kab. Pati (2), Kab. Banjarnegara (2), Kab. Batang (2), Kota Pekalongan (2), Kab. Pekalongan (2), Kab. Kudus (1), Kab. Blora (1), Kab. Wonosobo (1), Kab. Temanggung (1), Kab. Sukoharjo (1), Kab. Magelang (1)		
Sufism and Psychotherapy (TP)	Kota Semarang (35), Kab. Kendal (19), Kab. Pati (14), Kab. Demak (13), Kab. Jepara (12), Kab. Tegal (10), Kab. Brebes (9), Kab. Rembang (7), Kab. Pemasang (7), Kab. Batang (7), Kab. Grobogan (6), Kab. Kudus (4), Kab. Kebumen (4), Kab. Pekalongan (4), Kab. Semarang (3), Kab. Temanggung (3), Kab. Blora (3), Kab. Boyolali (3), Kota Tegal (2), Kota Pekalongan (2), Kab. Purworejo (2), Kab. Klaten (2), Kab. Sragen (1), Kab. Karanganyar (1), Kab. Wonosobo (1), Kab. Banjarnegara (1), Kab. Sukoharjo (1)	0, 1	0, 2

While this discussion focuses on FST and FUHUM, similar patterns were observed across other faculties as well. For instance, the Faculty of Education and Teacher Training (FITK) and the Faculty of Da’wah and Communication Science tend to attract students with mixed educational backgrounds, combining both secular and religious profiles. Meanwhile, the Faculty of Sharia, the Faculty of Economics and Islamic Business (FEBI), and the Faculty of Social and Political Sciences (FISIP) exhibit more varied regional intake and tend to be balanced in terms of gender.

3.5 Discussion

This study utilized K-Means and Fuzzy C-Means clustering to map the potential of prospective new students at UIN Walisongo Semarang based on key variables: gender, admission pathway, school type, study program, and region of origin. The analysis successfully identified three dominant clusters with distinct educational profiles. The Faculty of Science and Technology and the Faculty of Ushuluddin and Humanities served as contrasting case studies, illustrating how students' educational backgrounds influence their program preferences. A significant contribution of this research lies in the development of cluster-based student mapping, as presented in Section 3.4. The mapping visually demonstrated the geographic and institutional concentration of each cluster across different faculties. For example, Cluster 1 students—typically from general high schools—were predominantly mapped to science and technology programs. In contrast, Cluster 2, comprising graduates of Islamic schools, was aligned with programs in the religious and humanities faculties. This provides concrete evidence that clustering results are not only computational outputs but can be translated into actionable insights for institutional planning and targeted outreach.

Unlike previous studies that focused primarily on predicting student performance after admission or analyzing the behavior of active learners (e.g., Nabila et al., 2021), this research introduces a shift in perspective by focusing on the pre-admission phase. It positions itself as a diagnostic tool rather than a predictive one—designed to support institutional strategy formulation

rather than student-level intervention. Previous applications of clustering often served internal academic purposes (e.g., risk mitigation, dropout prediction), whereas this study directs clustering techniques toward external engagement and enrollment planning. Furthermore, while earlier studies typically used a single clustering method, this research provides a comparative analysis of K-Means and Fuzzy C-Means, reinforced by multiple validity checks (Elbow Method, Silhouette Score, FPC, and Xie-Beni Index). This combination enhances methodological depth, ensuring that clustering outputs are both computationally reliable and institutionally interpretable.

The novelty of this research lies in its comprehensive, data-driven mapping of applicants before enrollment and its application of comparative clustering techniques to explore educational diversity. This provides universities with a practical analytical tool for understanding student segments and developing targeted promotional strategies, thereby bridging data analytics with institutional decision-making in the context of higher education admissions. These findings equip higher education institutions with a data-driven framework to enhance their promotion and recruitment strategies. By aligning outreach efforts with actual applicant segmentation, universities can enhance the efficiency of communication, improve conversion rates, and make more informed decisions regarding faculty-level enrollment targets. This conclusion directly addresses the research problem by offering both analytical insights and institutional utility, thereby fulfilling the original research objective in a measurable and meaningful way.

4. Conclusion

This study identified three primary clusters of prospective new students at UIN Walisongo Semarang using K-Means and Fuzzy C-Means clustering methods. The analysis was based on key attributes, including gender, type of school, admission pathway, and region of origin. The findings revealed notable contrasts in student profiles, particularly between the Faculty of Science and Technology and the Faculty of Ushuluddin and Humanities. The combined use of both clustering algorithms offered a more comprehensive perspective and strengthened the validity of the results. These findings affirm that mapping prospective students can serve as a valuable tool for forecasting and enhancing future promotional and admission strategies.

The practical implication of this research lies in its ability to support data-driven decision-making, particularly in understanding and segmenting prospective student populations. Future studies may expand this model by incorporating additional variables, such as academic achievement or motivational factors, and applying it to broader datasets or across multiple institutions to test its robustness and generalizability.

5. References

- Abdullah, D., Susilo, S., Ahmar, A. S., Rusli, R., & Hidayat, R. (2022). The Application of K-Means Clustering for Province Clustering in Indonesia of The Risk of The COVID-19 Pandemic Based on COVID-19 Data. *Quality and Quantity*, 56(3), 1283–1291. <https://doi.org/10.1007/s11135-021-01176-w>.
- Achmad, F. S., & Widiastuti, I. (2024). Educator Recruitment and Selection Strategies in Private

Universities in Indonesia : Policies and Practices. *AL-ISHLAH: Jurnal Pendidikan*, 16(4), 4419–4429. <https://doi.org/10.35445/alishlah.v16i4.6000>.

- Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The K-Means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics (Switzerland)*, 9(8), 1–12. <https://doi.org/10.3390/electronics9081295>.
- Al-Alawi, M. K., Jaddoa, A., Cugley, J., & Hassanin, H. (2024). A Novel Enhanced SOC Estimation Method for Lithium-ion Battery Cells using Cluster-based LSTM Models and Centroid Proximity Selection. *Journal of Energy Storage*, 97, 1–15. <https://doi.org/10.1016/j.est.2024.112866>.
- Amartha, M. R. (2020). Penyebaran Mahasiswa Baru Menggunakan Metode Fuzzy C-Means Untuk Mencari Daerah Promosi Yang Potensial Distribution of New Students Using the Fuzzy C-Means Method To Look for Potential Promotional Areas. *Journal of Information Technology and Computer Science (INTECOMS)*, 3(2), 102-112. <https://doi.org/10.31539/intecom.v3i2.1524>.
- Ampadu, Y. B. (2023). Handling Big Data in Education: A Review of Educational Data Mining Techniques for Specific Educational Problems. *AI, Computer Science and Robotics Technology*, 2(April), 1–16. <https://doi.org/10.5772/acrt.17>.
- Chin, S., & Collins, J. E. (2024). Clustering Methods in Rheumatic and Musculoskeletal Diseases Research: An Educational Guide to Best Research Practices. *The Journal of Rheumatology*, 51(1), 1160–1168. <https://doi.org/10.3899/jrheum.2024-0519>.
- Chong, B. (2021). K-Means Clustering Algorithm: A Brief Review. *Academic Journal of Computing & Information Science*, 4(5), 37–40. <https://doi.org/10.25236/ajcis.2021.040506>.
- Dubey, A. K., Gupta, U., & Jain, S. (2018). Comparative Study of K-Means and Fuzzy C-Means Algorithms on The Breast Cancer Data. *International Journal on Advanced Science, Engineering and Information Technology*, 8(1), 18–29. <https://doi.org/10.18517/ijaseit.8.1.3490>.
- Et-Taleby, A., Boussetta, M., & Benslimane, M. (2020). Faults Detection for Photovoltaic Field Based on K-Means, Elbow, and Average Silhouette Techniques through the Segmentation of a Thermal Image. *International Journal of Photoenergy*, 2020, 1–7. <https://doi.org/10.1155/2020/6617597>.
- Gatea, A. N., & Al-asadi, H. A. A. (2024). Soft Clustering Techniques : An In-Depth Analysis of GMM and FCM Algorithms and Comparative Performance. *Journal of Basrah Researches (Sciences)*, 50(2), 223–239. <https://doi.org/10.56714/bjrs.50.2.19>.
- Gewers, F. L., Ferreira, G. R., De Arruda, H. F., Silva, F. N., Comin, C. H., Amancio, D. R., & Costa, L. D. F. (2021). Principal Component Analysis: A Natural Approach to Data Exploration. *ACM Computing Surveys*, 54(4), 1–33. <https://doi.org/10.1145/3447755>.
- Habsy, B. A., Zakirah, A., Rahmah, M. A., & Nafisah, C. A. (2023). Implementasi Teori Kognitif dan Konstruktivisme dalam Pembelajaran bagi Peserta Didik. *Tsaqofah*, 4(1), 326–342. <https://doi.org/10.58578/tsaqofah.v4i1.2182>.
- Hairani, H., Susilowati, D., Puji Lestari, I., Marzuki, K., & Mardedi, L. Z. A. (2022). Segmentasi Lokasi Promosi Penerimaan Mahasiswa Baru Menggunakan Metode RFM dan K-Means Clustering. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(2), 275–282. <https://doi.org/10.30812/matrik.v21i2.1542>.

- Han, H. (2023). Fuzzy Clustering Algorithm for University Students' Psychological Fitness and Performance Detection. *Heliyon*, 9(8), e18550. <https://doi.org/10.1016/j.heliyon.2023.e18550>.
- Hassan, A. A. H., Shah, W. M., Othman, M. F. I., & Hassan, H. A. H. (2020). Evaluate the Performance of K-Means and the fuzzy C-Means Algorithms to Formation Balanced Clusters in Wireless Sensor Networks. *International Journal of Electrical and Computer Engineering*, 10(2), 1515–1523. <https://doi.org/10.11591/ijece.v10i2.pp1515-1523>.
- Huang, Z. Y., Lin, S., Long, L. L., Cao, J. Y., Luo, F., Qin, W. C., ... Gregersen, H. (2020). Predicting the Morbidity of Chronic Obstructive Pulmonary Disease Based on Multiple Locally Weighted Linear Regression Model with K-Means Clustering. *International Journal of Medical Informatics*, 139(April), 104141. <https://doi.org/10.1016/j.ijmedinf.2020.104141>.
- Ikotun, A. M., Almutari, M. S., & Ezugwu, A. E. (2021). K-Means-Based Nature-Inspired Metaheuristic Algorithms for Automatic Data Clustering Problems: Recent Advances and Future Directions. *Applied Sciences (Switzerland)*, 11(23). <https://doi.org/10.3390/app112311246>.
- Jia, H., & Cheung, Y. M. (2018). Subspace Clustering of Categorical and Numerical Data with An Unknown Number of Clusters. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8), 3308–3325. <https://doi.org/10.1109/TNNLS.2017.2728138>.
- Krasnov, D., Davis, D., Malott, K., Chen, Y., Shi, X., & Wong, A. (2023). Fuzzy C-Means Clustering: A Review of Applications in Breast Cancer Detection. *Entropy*, 25(7), 1–14. <https://doi.org/10.3390/e25071021>.
- Kusnadi, K., & Ningrum, S. U. D. (2024). National Politics Dr. Soetomo for Education in Indonesia: An Alternative to Strengthen the Pillars of Education. *Jurnal Humanitas: Katalisator Perubahan Dan Inovator Pendidikan*, 10(2), 206–219. <https://doi.org/10.29408/jhm.v10i2.25169>.
- Lei, T., Jia, X., Zhang, Y., He, L., Meng, H., & Nandi, A. K. (2018). Significantly Fast and Robust Fuzzy C-Means Clustering Algorithm Based on Morphological Reconstruction and Membership Filtering. *IEEE Transactions on Fuzzy Systems*, 26(5), 3027–3041. <https://doi.org/10.1109/TFUZZ.2018.2796074>.
- Ling, L. S., & Weiling, C. T. (2025). Enhancing Segmentation: A Comparative Study of Clustering Methods. *IEEE Access*, 13(January). <https://doi.org/10.1109/ACCESS.2025.3550339>.
- Liu, H. (2023). Modern University Academy System: A Useful Exploration of Promoting the Cultivation of Innovative Abilities of College Students. *International Journal of New Developments in Education*, 5(22), 134–140. <https://doi.org/10.25236/ijnde.2023.052223>.
- Marbouti, F., Ulas, J., & Wang, C. H. (2021). Academic and Demographic Cluster Analysis of Engineering Student Success. *IEEE Transactions on Education*, 64(3), 261–266. <https://doi.org/10.1109/TE.2020.3036824>.
- Miraftabzadeh, S. M., Colombo, C. G., Longo, M., & Foadelli, F. (2023). K-Means and Alternative Clustering Methods in Modern Power Systems. *IEEE Access*, 11(November), 119596–119633. <https://doi.org/10.1109/ACCESS.2023.3327640>.
- Motlagh, O., Berry, A., & O'Neil, L. (2019). Clustering of Residential Electricity Customers using Load Time Series. *Applied Energy*, 237(August 2018), 11–24. <https://doi.org/10.1016/j.apenergy.2018.12.063>.
- Mughnyanti, M., Efendi, S., & Zarlis, M. (2020). Analysis of Determining Centroid Clustering X-

- Means Algorithm with Davies-Bouldin Index Evaluation. *IOP Conference Series: Materials Science and Engineering*, 725(1). <https://doi.org/10.1088/1757-899X/725/1/012128>.
- Nabila, S. P., Ulinnuha, N., & Yusuf, A. (2021). Model Prediksi Kelulusan Tepat Waktu dengan Metode Fuzzy C-Means dan K-Nearest Neighbors menggunakan Data Registrasi Mahasiswa. *Network Engineering Research Operation*, 6(1), 39. <https://doi.org/10.21107/nero.v6i1.199>.
- Pampuch, L. A., Negri, R. G., Loikith, P. C., & Bortolozzo, C. A. (2023). A Review on Clustering Methods for Climatology Analysis and Its Application over South America. *International Journal of Geosciences*, 14(09), 877–894. <https://doi.org/10.4236/ijg.2023.149047>.
- Patel, E., & Kushwaha, D. S. (2020). Clustering Cloud Workloads: K-Means vs Gaussian Mixture Model. *Procedia Computer Science*, 171(2019), 158–167. <https://doi.org/10.1016/j.procs.2020.04.017>.
- Qin, J., Hu, F., Liu, Y., Witherell, P., Wang, C. C. L., Rosen, D. W., ... Tang, Q. (2022). Research and Application of Machine Learning for Additive Manufacturing. *Additive Manufacturing*, 52(October 2021). <https://doi.org/10.1016/j.addma.2022.102691>.
- Qorib, F. (2024). Tantangan dan Peluang Kolaborasi antara Perguruan Tinggi dan Masyarakat dalam Program Pengabdian di Indonesia. *Journal of Indonesian Society Empowerment*, 2(2), 46–57. <https://doi.org/10.61105/jise.v2i2.119>.
- Rajkumar, K. V., Yesubabu, A., & Subrahmanyam, K. (2019). Fuzzy clustering and Fuzzy C-Means partition cluster analysis and validation studies on a subset of CiteScore dataset. *International Journal of Electrical and Computer Engineering*, 9(4), 2760–2770. <https://doi.org/10.11591/ijece.v9i4.pp2760-2770>.
- Ran, X., Zhou, X., Lei, M., Tepsan, W., & Deng, W. (2021). A Novel K-Means Clustering Algorithm with a Noise Algorithm for Capturing Urban Hotspots. *Applied Sciences*, 11(23). <https://doi.org/10.3390/app112311202>.
- Rayala, V., & Kalli, S. R. (2021). Big Data Clustering using Improvised Fuzzy C-Means Clustering. *Revue d'Intelligence Artificielle*, 34(6), 701–708. <https://doi.org/10.18280/RIA.340604>.
- Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. da F., & Rodrigues, F. A. (2019). Clustering Algorithms: A Comparative Approach. *In PLoS ONE*, 14. <https://doi.org/10.1371/journal.pone.0210236>.
- Sadeghi, M., Casey, P., Carranza, E. J. M., & Lynch, E. P. (2024). Principal Components Analysis and K-Means Clustering of Till Geochemical Data: Mapping and Targeting of Prospective Areas for Lithium Exploration in Västernorrland Region, Sweden. *Ore Geology Reviews*, 167(January), 106002. <https://doi.org/10.1016/j.oregeorev.2024.106002>.
- Sahin, A. (2018). Critical Issues in Islamic Education Studies: Rethinking Islamic and Western Liberal Secular Values of Education. *Religions*, 9(11). <https://doi.org/10.3390/rel9110335>.
- Sammouda, R., & El-Zaart, A. (2021). An Optimized Approach for Prostate Image Segmentation Using K-Means Clustering Algorithm with Elbow Method. *Computational Intelligence and Neuroscience*, 2021. <https://doi.org/10.1155/2021/4553832>.
- Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2021). A Quantitative Discriminant Method of Elbow Point for The Optimal Number of Clusters in Clustering Algorithm. *Eurasip Journal on Wireless Communications and Networking*, 2021(1). <https://doi.org/10.1186/s13638-021->

01910-w.

- Singh, J., & Singh, D. (2024). A Comprehensive Review of Clustering Techniques in Artificial Intelligence for Knowledge Discovery: Taxonomy, Challenges, Applications and Future Prospects. *Advanced Engineering Informatics*, 62. <https://doi.org/10.1016/j.aei.2024.102799>.
- Yin, Y., Yong, Y., Qi, S., Yang, K., & Lan, Y. (2023). Cluster Analyses of Tropical Cyclones with Genesis in the South China Sea Based on K-Means Method. *Asia-Pacific Journal of Atmospheric Sciences*, 59(4), 433–446. <https://doi.org/10.1007/s13143-023-00322-8>.
- Zaib, R., & Ourabah, O. (2023). Large Scale Data Using K-Means. *Mesopotamian Journal of Big Data*, 2023(2023), 36–45. <https://doi.org/10.58496/mjbd/2023/006>.
- Žalėnienė, I., & Pereira, P. (2021). Higher Education For Sustainability: A Global Perspective. *Geography and Sustainability*, 2(2), 99–106. <https://doi.org/10.1016/j.geosus.2021.05.001>.
- Zarkasyi, M. R. (2024). Organizational Commitment of the Millennial Generation Workforce in Indonesian Higher Education: Systematic Literature Review. *AL-ISHLAH: Jurnal Pendidikan*, 16(2), 1097–1107. <https://doi.org/10.35445/alishlah.v16i2.5151>.
- Zhang, Y., Wu, X., Xu, J., Ning, Z., & Han, X. (2025). Clustering Analysis of Yue Opera Character Tone Trends Based on Quantum Particle Swarm Optimization for Fuzzy C-Means. *PLoS ONE*, 20(1 January), 1–25. <https://doi.org/10.1371/journal.pone.0313065>.
- Zhou, S., Li, D., Zhang, Z., & Ping, R. (2021). A New Membership Scaling Fuzzy C-Means Clustering Algorithm. *IEEE Transactions on Fuzzy Systems*, 29(9), 2810–2818. <https://doi.org/10.1109/TFUZZ.2020.3003441>.