

## STUDENT ACADEMIC ACHIEVEMENT CLUSTERING USING FUZZY C-MEANS ALGORITHM

Natria Selina<sup>1\*</sup>, Sriani<sup>1</sup>

<sup>1</sup>Computer Science, Universitas Islam Negeri Sumatera Utara

*email: \*natriaselina06@gmail.com*

**Abstract:** Academic achievement mapping is an important process in higher education to support effective academic monitoring and guidance. In practice, student grouping is often conducted manually by academic staff using simple criteria such as Grade Point Average (GPA) thresholds and subjective judgment, without systematic data analysis. This study aims to apply the Fuzzy C-Means (FCM) clustering algorithm to objectively group students based on their academic achievement levels. The dataset consists of academic records from 179 sixth-semester students of the Computer Science Study Program at Universitas Islam Negeri Sumatera Utara, where 160 eligible students are processed in the FCM calculation. Three variables are used: cumulative GPA, total completed credits, and the total number of low grades (D/E). The FCM algorithm automatically performs the mapping and groups students into three categories, namely excellent, stable, and at-risk students. Cluster quality is evaluated using the Silhouette Score and Davies–Bouldin Index, showing satisfactory clustering performance. The results indicate that the proposed approach provides a data-driven and objective basis for academic decision support.

**Keywords:** academic achievement; clustering; fuzzy c-means; student

**Abstrak:** Pemetaan pencapaian akademik mahasiswa merupakan proses penting dalam pendidikan tinggi untuk mendukung pemantauan dan pembinaan akademik yang tepat sasaran. Dalam praktiknya, pengelompokan mahasiswa masih sering dilakukan secara manual oleh pihak akademik berdasarkan kriteria sederhana, seperti batasan Indeks Prestasi Kumulatif (IPK) dan penilaian subjektif, tanpa analisis data yang sistematis. Penelitian ini bertujuan menerapkan algoritma Fuzzy C-Means (FCM) untuk mengelompokkan mahasiswa secara objektif berdasarkan tingkat pencapaian akademik. Data penelitian berasal dari 179 mahasiswa semester enam Program Studi Ilmu Komputer Universitas Islam Negeri Sumatera Utara, dengan 160 mahasiswa memenuhi kriteria dan diproses menggunakan algoritma FCM. Variabel yang digunakan meliputi IPK kumulatif, jumlah SKS yang telah ditempuh, dan total nilai rendah (D/E). Proses pemetaan sepenuhnya dilakukan oleh algoritma FCM dan menghasilkan tiga kategori mahasiswa, yaitu unggul, stabil, dan berisiko. Evaluasi menggunakan Silhouette Score dan Davies–Bouldin Index menunjukkan kualitas pengelompokan yang cukup baik.

**Kata kunci:** fuzzy c-means; clustering; mahasiswa; pencapaian akademik

## INTRODUCTION

Student academic achievement is one of the main indicators used to evaluate the effectiveness of the learning process in higher education institutions [1]. This achievement is commonly represented by indicators such as Grade Point Average (GPA), the number of completed credits, and the consistency of students' academic performance [2], [3]. Information on academic achievement plays a strategic role for institutions in evaluating learning outcomes, formulating academic policies, and designing sustainable student development programs [4].

Along with the increasing number of students and the growing complexity of academic data, manual evaluation and grouping processes have become less effective and are prone to subjectivity [5], [6]. Although academic information systems store large volumes of student data, these data are often used only for administrative reporting and have not been optimally analyzed to generate strategic insights for decision making [7]. This condition highlights the need for data-driven analytical approaches that can process academic data objectively, accurately, and efficiently [8], [9].

The application of data mining techniques in the educational domain, known as educational data mining, has been widely adopted to extract meaningful patterns and knowledge from academic data [10]. One of the commonly used techniques is clustering, which groups data based on similarity among certain characteristics [11], [12]. Through clustering, students can be classified into several academic categories that represent their achievement levels, thereby supporting academic monitoring and decision-making processes [13].

Several previous studies have demonstrated the effectiveness of clustering techniques in analyzing student academic performance. For instance, prior research applied K-Means clustering to classify students based on GPA and course completion patterns, successfully identifying groups of high-risk and high-performing students [13]. Other studies utilized hierarchical clustering to explore learning behavior patterns and academic progression trends, providing insights for early academic intervention programs. Additionally, comparative studies between hard clustering and soft clustering approaches indicate that soft clustering methods are more flexible in representing complex student characteristics, as students may exhibit overlapping academic attributes [8].

Fuzzy C-Means (FCM) extends conventional hard clustering by permitting data points to have partial membership in more than one cluster [14], [15], [16]. This characteristic makes FCM particularly suitable for modeling student academic conditions, which are often uncertain and cannot be strictly classified into a single category [17]. Although previous studies have applied FCM in educational contexts, its application for grouping students based on academic indicators as a foundation for academic decision support systems remains limited, especially in different institutional and data contexts [18].

Some recent studies have implemented FCM to evaluate academic risk levels and student learning outcomes, showing improved interpretability of cluster structures compared to conventional partitioning methods [19]. However, most of these studies focus only on classification results without integrating the clustering outcomes into a structured academic decision support framework.

Furthermore, variations in institutional data characteristics—such as differences in curriculum structures, credit systems, and grading policies—suggest the need for contextualized implementation and validation of FCM models [20].

Based on this research gap, this study aims to apply the Fuzzy C-Means algorithm to group students according to their academic achievement levels using indicators such as GPA, completed credits, and other relevant academic data. The resulting clusters are expected to provide an objective overview of students’ academic conditions and serve as a basis for developing an academic decision support system to assist institutions in designing more targeted and data-driven student guidance and intervention strategies [21].

**METHOD**

This study adopts a data mining approach using clustering techniques to group students based on their academic achievement levels. The research subjects are sixth-semester undergraduate students of the Computer Science Study Program at Universitas Islam Negeri Sumatera Utara (UINSU) Medan, class of 2022. The dataset consists of 179 student records, of which 160 students meet the data completeness criteria and are processed using the Fuzzy C-Means (FCM) algorithm, while the remaining data are excluded due to incomplete academic records. The dataset includes cumulative Grade Point Average (GPA) from semesters 1–6, the total number of completed credits, and the total number of low grades (D/E) as indicators of academic risk. All data were obtained from the faculty academic database with due consideration of research ethics and data confidentiality [22].

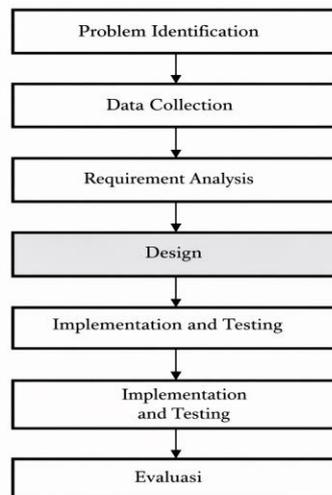


Figure 1. Research Framework

The research framework begins with identifying problems related to the lack of objective and data-driven student grouping. The research stages include data collection and preprocessing, determination of the number of clusters, and application of the Fuzzy C-Means algorithm through iterative processes until convergence is achieved. The clustering results are used to classify students into three categories: excellent, stable, and at-risk students. Cluster quality is evaluated using the Silhouette Coefficient and Davies–Bouldin Index to assess cluster separation and compactness. The results are presented in the form of tables, graphs, and cluster visualizations as a basis for developing an academic decision support system [23].

**RESULT AND DISCUSSION**

**Data Analysis and Dataset**

This study developed a student grouping system based on academic achievement using the Fuzzy C-Means (FCM) algorithm. The dataset was obtained from the academic transcript records (Kartu Hasil Studi/KHS) of students

from semesters 1–6 in the Undergraduate Computer Science Program at Universitas Islam Negeri Sumatera Utara (UIN-SU) Medan, class of 2022, with a total of 179 students. Of these, 160 students met the data completeness criteria ( $\geq 6$  semesters) and were processed using FCM, while 19 students were not processed due to incomplete semester data.

The variables used include cumulative GPA from semesters 1–6, total completed credits (SKS) from semesters 1–6, and total number of D/E grades from semesters 1–6, representing academic performance, academic progress, and academic risk, respectively.

Table 1. Summary of the Research Dataset (Key Information)

Component	Number
Total student data	179
Eligible data ( $\geq 6$ semesters)	160
Incomplete data ( $< 6$ semesters)	19

### Data Preprocessing and Normalization

The preprocessing stage included data completeness validation, numeric data type conversion, duplicate removal based on student ID (NIM), and elimination of missing values in the main features. Subsequently, the eligible data were normalized using the Min–Max

### Cluster Mapping and Interpretation

Based on centroid characteristics, each cluster was mapped to an academic category as follows:

Table 4. Cluster-to-Category Mapping

Cluster	Category
0	At-Risk
1	Stable
2	Excellent

normalization method to equalize feature scales prior to the clustering process.

Table 2. Minimum and Maximum Feature Values (Eligible Data)

Feature	Minimum	Maximum
GPA semesters 1–6	1.48	3.83
Total credits (SKS) semesters 1–6	89	136
Total D/E grades semesters 1–6	0	21

### Implementation of Fuzzy C-Means

The FCM algorithm was applied with three clusters ( $c = 3$ ) and a fuzziness coefficient ( $w = 2$ ). The iterative process continued until convergence was achieved, resulting in the final cluster centroids and membership degrees for each student. Based on the centroid patterns, the clusters were interpreted in an academic context.

Table 3. Final Cluster Centroids (Original Scale)

Cluster	GPA	Credits (SKS)	Total D/E
0	1.93	106.04	16.75
1	3.12	124.76	4.54
2	3.50	128.61	0.56

The Excellent cluster is characterized by high GPA and credit completion with a low number of D/E grades, indicating strong academic performance and the potential for early graduation. The Stable cluster reflects moderate academic achievement with a high likelihood of on-time graduation. Meanwhile, the At-Risk cluster shows low GPA and credit completion with a high number of D/E grades, indicating potential delays in graduation.

**Student Grouping Results**

The distribution of clustering results was analyzed for both the entire dataset and the eligible data subset.

Table 5. Student Grouping Summary (Total 179 Students)

Category	Number	Percentage (%)
Excellent	69	38.55
Stable	82	45.81
At-Risk	9	5.03
Not processed	19	10.61

Table 6. Student Grouping Summary (Eligible Data – 160 Students)

Category	Number	Percentage (%)
Excellent	69	43.12
Stable	82	51.25
At-Risk	9	5.62

The results indicate that the Stable category dominates, followed by Excellent, while the At-Risk category has the smallest proportion but still requires academic attention.

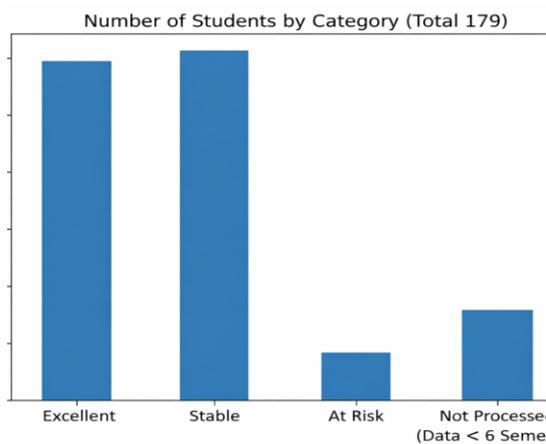


Figure 3. Number of Students per Category (Total 179 Students)

Percentage by Category (Eligible only)

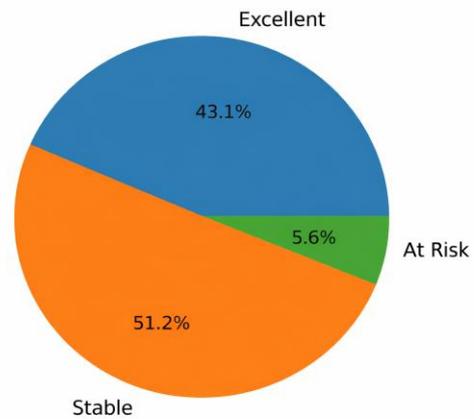


Figure 4. Percentage Distribution of Categories (Eligible Data – 160 Students)

**Clustering Quality Evaluation**

The quality of the clustering results was evaluated using internal validation metrics.

Table 7. Clustering Evaluation Results

Metric	Value	Interpretation
Silhouette Score	0.5535	Adequate cluster separation
Davies–Bouldin Index	0.6230	Relatively compact cluster structure

A Silhouette Score above 0.5 and a relatively low DBI value indicate that the clustering results are sufficiently reliable for interpretation and decision-making.

**System Output**

The system produces final outputs in Excel (.xlsx) format containing student data, clustering results, membership degrees, and final academic categories (Excellent, Stable, At-Risk, and Not Processed). These outputs are ready to be

used as a basis for academic reporting and decision support systems.

## CONCLUSION

This study demonstrates the effective application of the Fuzzy C-Means (FCM) algorithm for grouping students based on academic achievement from semesters 1–6 using cumulative GPA, total completed credits, and the number of D/E grades as key indicators. The proposed system successfully performs data preprocessing, Min–Max normalization, and iterative clustering to generate meaningful membership values and cluster centroids, which are subsequently mapped into three academic categories: Excellent, Stable, and At-Risk. The clustering results show that the Stable category dominates, followed by Excellent, while At-Risk students represent a smaller proportion but require targeted academic intervention. Cluster quality evaluation produced a Silhouette Score of 0.5535 and a Davies–Bouldin Index of 0.6230, indicating a sufficiently well-separated and compact clustering structure. Moreover, the system accommodates students with incomplete semester records by including them in the final output, supporting comprehensive academic reporting. Future work should consider additional academic variables, alternative clustering techniques, larger datasets, and the development of a fully integrated academic monitoring and decision support system.

## BIBLIOGRAPHY

[1] R. Ghaniy and F. Indriyaningsih, “Penerapan Metode Fuzzy C-

Means dalam Pemilihan Program Studi Mahasiswa Baru di Perguruan Tinggi,” vol. 10, no. 2, pp. 19–30, 2020.

[2] H. Jamhur, “Pemodelan Prediksi Predikat Kelulusan Mahasiswa menggunakan Fuzzy C-Means berbasis Particle Swarm Optimizin,” vol. 10, no. 1, pp. 13–24, 2020.

[3] R. Nurul, J. S. Intam, A. Akram, N. Risal, and D. F. Suroianto, “Klasifikasi Mahasiswa Berprestasi Menggunakan Fuzzy C-Means Dan Naive Bayes,” vol. 15, no. 1, pp. 9–16, 2024.

[4] M. Anugraha, A. Eka, A. Hartanto, and R. P. Kristianto, “KLAUSTERISASI PENGELOLAAN BAN DI PT . GST MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING,” vol. 3, no. 2, pp. 312–326, 2024.

[5] S. Syaquila and M. Fakhriza, “K-Means Clustering Untuk Mengukur Pengaruh Kompetensi Terhadap Kinerja Pegawai,” vol. 6, no. 2, pp. 1420–1431, 2025, doi: 10.47065/josh.v6i2.6758.

[6] D. Rizki, P. Noordi, I. A. Prastowo, N. A. Putri, Z. A. Ekovich, and D. Hartanti, “Penerapan Algoritma Fuzzy C-Means untuk Analisis Tingkat Pendidikan Sekolah Dasar di Karanganyar,” pp. 12–18, 2023.

[7] F. Hariyadi, H. Sujaini, and A. S. S, “Pengklausteran Mahasiswa UNTAN Berdasarkan IPK Menggunakan Metode Fuzzy C-Means UNTAN Student Clustering Based on IPK Using Fuzzy C-Means Method,” pp. 31–37, 2020, doi: 10.26418/justin.v8i2.36506.

- [8] N. Fadiah, R. Nurul, J. S. Intam, and D. Fatmarani, "Analisis Penentuan Mahasiswa Berprestasi Fakultas Teknik UNM Menggunakan Metode Fuzzy C-Means Analysis of the Determination of Outstanding Students of the Faculty of Engineering UNM Using the Fuzzy C-Means Method," vol. 11, no. 3, pp. 480–486, 2023, doi: 10.26418/justin.v11i3.
- [9] R. Sovia, E. Praja, W. Mandala, and S. Mardhiah, "Algoritma K-Means dalam Pemilihan Siswa Berprestasi dan Metode SAW untuk Prediksi," vol. 6, no. 2, pp. 181–187, 2020.
- [10] G. Nabila, S. Putri, D. Ispriyanti, T. Widiharah, D. Statistika, and U. Diponegoro, "Implementasi algoritma fuzzy c-means dan fuzzy possibilistics c-means untuk klasterisasi data tweets pada akun twitter tokopedia 1,2,3," vol. 11, pp. 86–98, 2022.
- [11] M. R. Nahjan, N. Heryana, A. Voutama, F. I. Komputer, U. S. Karawang, and R. Miner, "IMPLEMENTASI RAPIDMINER DENGAN METODE CLUSTERING K-MEANS UNTUK ANALISA PENJUALAN PADA TOKO OJ CELL," vol. 7, no. 1, pp. 101–104, 2023.
- [12] M. R. Gusmansyah, S. Daulay, and A. Rivaldi, "Perbandingan Metode Machine Learning untuk Klastering Penerima Bantuan Pendidikan Siswa," pp. 2–8, 2025.
- [13] S. Agustinus, B. Telaumbanua, F. Setiadi, and S. Nurjanah, "Analisis Clustering Menggunakan Metode Enhanced Fuzzy C-Means Clustering Dengan Algoritma Rock Pada Student Performance Dataset," vol. 7, no. 3, 2025, doi: 10.32877/bt.v7i3.2287.
- [14] H. K. Candra, M. Bahit, B. Sabella, K. Akuntansi, and P. N. Banjarmasin, "PENERAPAN METODE KLUSTERING FUZZY C-MEANS UNTUK PENENTUAN PEMINATAN PEMILIHAN JURUSAN Kata," 2021.
- [15] Mirand and Sriani, "View of Implementation of K-Means Clustering in Grouping Sales Data at Zura Mart.pdf," 2025.
- [16] F. Khalish, N. M. Piranti, and O. Martdireja, "View of Implementasi Data Mining Menggunakan Teknik Clustering dengan Metode K-Means," 2025.
- [17] R. A. CHANIAGO, *MENULAR DI KOTA SIBOLGA MENGGUNAKAN ALGORITMA FUZZY C-MEANS SKRIPSI Disusun Sebagai Syarat Memperoleh Gelar Sarjana Komputer Prodi Teknik Informatika Fakultas Teknik Universitas Malikussaleh DISUSUN OLEH: NAMA PRODI: REZA ALFIAN CHANIAGO JURUSAN INFORMA.* 2024.
- [18] F. Wilyani, Q. N. Arif, and F. Aslimar, "Pengenalan Dasar Pemrograman Python Dengan Google Colaboratory Basic," no. 1, 2024.
- [19] T. Zatul and H. Tengku, "Predicting Student Performance Using Subtractive Clustering Fuzzy C-Means and Multiple Linear Regression," vol. 16, no. 1, pp. 31469–31474, 2026.
- [20] D. N. Aini, "Fuzzy Clustering-Based Grouping for Mapping the

- Distribution of Student Success Data,” vol. 3, no. October, pp. 366–372, 2023.
- [21] D. K. D. Damayanti, “KLASIFIKASI STATUS STUNTING BALITA MENGGUNAKAN ALGORITMA FUZZY C-MEANS (STUDI KASUS POSYANDU RW 01 KELURAHAN JEPARA SURABAYA,” vol. 11, no. 03, pp. 524–533, 2023.
- [22] E. O. Srirahmawati *et al.*, “PENGELOMPOKAN PRESTASI AKADEMIK SISWA SD MENGGUNAKAN ALGORITMA K-MEANS,” vol. 8, no. 1, pp. 80–86, 2025.
- [23] S. Mujilawati and R. Wardhani, “IMPLEMENTASI FUZZY C-MEANS UNTUK CLUSTERING MAHASISWA BERDASARKAN NILAI MASUK PERGURUAN TINGGI,” vol. 6, no. 1, pp. 448–453, 2021.