



Comparison of K-Means Clustering and Fuzzy Tsukamoto Algorithms for Grouping Student Data of NU Medan Middle School Based on Academic Achievement

Abdul Rasyid Pahlevi¹, Amrullah²

^{1,2} Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

Article Info

Article history

Received : Oct 04, 2025

Revised : Oct 23, 2025

Accepted : Oct 25, 2025

Keywords:

Clustering;
Educational Data Analysis;
Fuzzy Tsukamoto;
K-Means;
Student Achievement.

Abstract

The increasing need for data-driven decision-making in education has encouraged the use of intelligent algorithms to evaluate and classify student academic performance more effectively. However, differences in algorithmic approaches often lead to variations in interpretation and categorization outcomes. This study aims to compare the performance of the K-Means and Fuzzy Tsukamoto algorithms in clustering student achievement data at SMP NU Medan to determine which method provides a more accurate and interpretable classification model. The research employs quantitative analysis using students' semester grades processed through Python and Microsoft Excel, where K-Means utilizes centroid-based clustering (75, 85, and 95) and Fuzzy Tsukamoto applies fuzzy logic with weighted membership values (0, 5, and 10). The results reveal that K-Means produces a more proportional and stable clustering structure, effectively differentiating student achievement levels within the same population, while Fuzzy Tsukamoto offers a simpler, rule-based classification system aligned with fixed academic standards. The findings indicate that K-Means is more suitable for analyzing relative performance variations, whereas Fuzzy Tsukamoto is better suited for absolute classification and administrative evaluation. Both methods are easily implemented and can be integrated into educational management systems to enhance instructional decision-making. The study implies that a hybrid combination of these two algorithms may provide a more comprehensive analytical framework for evaluating student performance.

Corresponding Author:

Abdul Rasyid Pahlevi
Teknologi Informasi,
Universitas Muhammadiyah Sumatera Utara
Jl Kapt Mukhtar Basri No 3 Glugur Darat II, Medan, 20239. Indonesia
Email : rasyidabdul453@gmail.com

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



1. Introduction

Education plays a central role in developing high-quality human resources and shaping a nation's competitiveness in the global era. Within formal education systems, student academic achievement serves as a key indicator of the success of learning processes and the overall quality of education in an institution (Tilak, 2021). Academic achievement reflects not only students' cognitive capacities but also their affective and psychomotor development, which together form a comprehensive

measure of learning outcomes (Anderson & Krathwohl, 2019). Effective analysis of academic performance enables educational institutions to make evidence-based decisions that enhance curriculum design, learning methods, and student support strategies (Han & Ellis, 2020). SMP NU Medan, as a developing junior high school in Indonesia, faces challenges in managing increasingly complex academic data due to the rising number of students each year. The accumulation of academic data across subjects and semesters requires an analytical model capable of identifying meaningful patterns (Kurniawan et al., 2022). However, most schools still rely on manual approaches or descriptive statistics that cannot capture deeper insights (Riza & Putri, 2021). To improve educational management, it is necessary to integrate information technology and data-driven techniques into academic evaluation systems (Suhendar & Rahman, 2020). The use of data mining in educational contexts—often referred to as Educational Data Mining (EDM)—has been widely recognized as a promising approach for analyzing large datasets and uncovering hidden knowledge in student performance data (Romero & Ventura, 2020; Baker et al., 2021). EDM allows institutions to transform raw academic data into actionable insights that support strategic planning and decision-making (Dutt et al., 2019).

Among various data mining techniques, clustering has emerged as one of the most useful for grouping data based on similarities in characteristics (Jain, 2010). Clustering is categorized as an unsupervised learning approach because it does not rely on predefined class labels but instead identifies structures or patterns that naturally exist in the data (Han et al., 2022). One of the most popular clustering algorithms is K-Means, known for its simplicity, computational efficiency, and ability to handle large datasets (MacQueen, 1967; Hartigan & Wong, 1979). The algorithm partitions data into k clusters by minimizing the distance between data points and the centroid of each cluster (Lloyd, 1982). In the context of education, K-Means can be used to group students based on academic performance indicators, thereby helping teachers identify students who excel or require additional support (Agustin et al., 2021). The results of clustering can serve as the basis for designing personalized learning strategies tailored to students' specific learning needs (Sembiring et al., 2020). However, despite its effectiveness, the K-Means algorithm assumes that each data point belongs exclusively to one cluster, which may not fully reflect the complexity of human academic performance (Wagstaff et al., 2001). In reality, students' abilities are often continuous and overlapping—some students may exhibit characteristics that lie between “medium” and “high” performance categories (Rasmani & Shen, 2006). This limitation suggests the need for an alternative or complementary method capable of modeling such uncertainty in data interpretation (Kusumadewi & Hartati, 2010).

Previous studies have demonstrated the potential of K-Means in analyzing educational datasets, yet they also highlight several research gaps. For example, Agustin, Kurniawati, and Heriyanto (2022) applied the K-Means algorithm to determine enrichment groups for national exam subjects, finding that the method effectively grouped students based on performance levels. However, their study was limited to national exam scores and did not incorporate subject-level data. Similarly, Priyatman et al. (2020) utilized K-Means for predicting student graduation times using GPA data, while Dacwanda et al. (2021) implemented K-Means to analyze knowledge and skills values, forming three clusters—smart, average, and sufficient—with 10.15% inter-cluster transition. Although informative, these studies largely focus on aggregated data and neglect contextual variations across individual subjects. Moreover, K-Means's hard partitioning structure prevents it from capturing the subtle gradations of academic performance (Chen et al., 2021). In contrast, fuzzy logic-based methods such as Fuzzy Tsukamoto offer an alternative that incorporates degrees of membership, allowing data points to belong partially to multiple clusters (Tsukamoto, 1979). The Fuzzy Tsukamoto method models linguistic variables and applies a rule-based system with monotonic membership functions to produce interpretable results (Kusumadewi & Purnomo, 2013). Therefore, combining K-Means and Fuzzy Tsukamoto in a comparative study can potentially provide more comprehensive insights into the distribution and dynamics of student achievement (Setiawan et al., 2022).

This study aims to address the identified research gaps by developing a comparative analytical model that integrates K-Means clustering and the Fuzzy Tsukamoto method to analyze academic achievement data of students at SMP NU Medan. The main research questions include: (1) How can

student academic data be effectively grouped using K-Means and Fuzzy Tsukamoto algorithms? (2) Which method yields more accurate and interpretable results for educational decision-making? and (3) How can the clustering results contribute to improving the quality of learning and educational management at SMP NU Medan? The objectives of this research are to (a) implement and compare the two methods on subject-level academic data, (b) evaluate the accuracy and interpretability of clustering outcomes, and (c) provide actionable recommendations for teachers and administrators in enhancing educational strategies. The study's novelty lies in its contextual focus on SMP NU Medan and its subject-level data analysis, rather than aggregated GPA, as well as in its use of a soft clustering approach (Fuzzy Tsukamoto) to handle uncertainty in student performance categorization (Huda et al., 2023). By integrating quantitative data analysis with pedagogical interpretation, this study aims to bridge the gap between data-driven insights and educational practice (Romero & Ventura, 2020; Hidayat et al., 2022).

Conceptually, the research follows a logical flow that begins with the collection of students' academic achievement data per subject, which is then processed using both K-Means and Fuzzy Tsukamoto algorithms. The results of both clustering methods are compared in terms of accuracy, interpretability, and their implications for educational management. Through this comparative framework, the study seeks to demonstrate how data mining techniques can transform raw academic data into actionable insights that support evidence-based educational decisions (Witten et al., 2017). The integration of fuzzy logic allows for a more flexible and realistic representation of students' performance levels, accommodating the inherent uncertainty in academic assessment (Zadeh, 1996). This approach is expected to assist schools in identifying student groups requiring intervention, designing adaptive learning programs, and optimizing resource allocation (Sembiring et al., 2020; Kusumadewi & Hartati, 2010). Ultimately, the contribution of this research lies in providing a practical and methodological framework for schools to leverage data mining techniques—particularly the integration of K-Means and Fuzzy Tsukamoto—in improving the quality of educational decision-making and supporting continuous learning improvement (Han et al., 2022; Baker et al., 2021).

2. Research Methodology

The research phase in comparing the algorithmic K-Means clustering in Fuzzy Tsukamoto to group students' data based on academic achievement is carried out by collecting data on the values of SMP NU Medan students from semester 1 to semester 4, which are then processed in preprocessing stages such as data cleaning and calculating the values of each student's calculations. This research aims to compare the performance of the algorithmic K-Means Clustering in Fuzzy Tsukamoto in grouping students' data based on academic achievement. The research flow includes:

1. Data Collection
Data for NU Medan Middle School students is collected from semester 1 to semester 4. Data includes the values of each study group, which will become input variables for subsequent processing.
2. Data Preprocessing
This process includes:
 - a. Data cleaning to remove blank values, duplicates, and invalid data.
 - b. Calculation of the values per study group for each student.
 - c. Data normalization to normalize the attribute values so that no valid values dominate the processing process.
3. Determining Criteria & Parameters
 - a. Determine the number of clusters according to the grouping objective, generally three categories: High Achievement, Medium Achievement, and Low Achievement.
 - b. Determine algorithm parameters, such as the centroid value for K-Means or the Fuzzy Tsukamoto membership function.
4. Process with K-Means Algorithm
 - a. Initialize the initial centroid points.

- b. Caculate the path of each point to al centroids.
- c. Group the points into clusters using the closest path
- d. Baance the centroid positions based on the cluster member vaues.
- e. Repeat until the centroid positions are stable (converged).
- f. Repeat the process until the cluster is baanced.
5. Process with Fuzzy Tsukamoto Algorithm
 - a. Fuzzy caculation: find the vaues of the student vaues into linguistic sets (Low, Middle, High).
 - b. Baance the IF-THEN equations that have been determined.
 - c. Inference: Caculate the predicate al-in vauue (μ) for each equation and determine the crisp parsimony/z-vauue (α).
 - d. Defuzzification using weighted ratios to produce the fina performance score.
 - e. Assign scores to performance categories according to the baance equations.
6. Hasil Anaysis & Comparison
 - a. Baancing the K-Means clustering equation in Fuzzy Logica Modeling using evauation metrics such as the Silhouette Coefficient, Cohen's Kapa, and the Al-Adjusted Rand Index.
 - b. Interpretasikan perbedaan hasil untuk melijat kelarian dan kelemahan masing-masing methods dalam teksta siswa.

K-Means Stages

1. Determining the Number of Clusters (K)
First, the researcher determined the number of clusters (K) used in the interna grouping of NU Medan Middle School students. Based on the research objective of grouping students based on academic achievement, three clusters were selected: students with high, medium, and low academic achievement. The selection of the $K = 3$ vauue was based on pedagogica considerations in the genera pattern of academic classification used in schools, so that the interna grouping results would be easier for school officias to interpret.
2. Determining the Centroid
After the number of clusters is determined, the next step is to select three centroids from the student vaues of NU Medan Middle School. The vaues used include core subject vaues such as Mathematics, Indonesian Language, English Language, and IPA from semesters 1 to 4. Each centroid represents a student group with certain vauue characteristics, which then become a key element in the clustering process.
3. Caculating Each Student's Path to the Centroid
At this stage, each student's path is caculated by comparing the centroids using the Euclidean Distance formula. For example, if a student has high grades in al subjects, their path is likely to be closer to the centroid of the high-achieving group. This process is carried out for al students at NU Medan Middle School, so that each student's path is identified using the predetermined clusters.
2. Grouping Students into the Nearest Cluster
4. Based on the path caculation, each student at NU Medan Middle School is then grouped into the cluster that has the path closest to the centroid vauue. For example, students with consistently high academic achievement scores fal into the high achievement cluster, while students with fluctuating low academic achievement scores fal into the low achievement cluster. This grouping provides a genera overview of the distribution of academic achievement within the school environment.
5. Recaculating the Centroid of Each Cluster
After the initia grouping is complete, the centroid of each cluster is recaculated. This process involves caculating the relative vaues of al students in each cluster. For example, a cluster with many high-achieving students will produce a new centroid with high relative vaues. This process is important to adjust the centroid position to be more representative of the group members at SMP NU Medan.
6. Repeating the Process Until No Cluster Changes

The fina step is to repeat the process of caculating paths, grouping, and changing centroids until no more students change clusters, and with each cluster, the grouping process is stable. In the context of NU Medan Middle School students, this process is carried out until al students remain in the same achievement group over severa consecutive iterations. This fina step is then used as an anaysis to see the distribution of students based on the achievement category.

3. Results and Discussion

The implementation of the K-Means algorithm for the grade IX students resulted in grouping into three categories, namely Low, Medium, and High, according to the nearest centroid (75, 85, and 95). The caculation process was carried out by caculating the ratios of each student from semester I to IV, then the ratios became the fina vaue. This fina vaue is used as a basis for determining the cluster. From the caculation results, it can be seen that most of the students in the Middle category, with their vaues centered around the centroid of 85. The grouping results show that the High category is filled by students with vaues around the centroid of 87, such as Syarifah Diza Al-Priyani, Sri Rindiyani Hafizai, and Al-Priyani Kartini. Meanwhile, the Low category is filled by students with vaues around the centroid of 80, such as Zais Revan Al-Bimanyu, Nurul Cahya Ramadhani, and Safira Mulhimai. Adapun students who are not completely, such as Radit Arya Pranata Saragih, are automaticaly categorized as “Not completely” and not completely mapped into clusters. Thus, the implementation of the K-Means program is able to provide a more objective picture of the distribution of student achievement based on the vaues of the Average.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	NAMA SISWA	PAI_AVG	PPKN_AVG	IDONESIA	TEMATIKA	IPA_AVG	IPS_AVG	INGGRIS	BUDAYA	PJOK_AVG	AKARYA	AIQRA_AVG	TIK_AVG	RATA_RATA	KATEGORI	
2	Abdul Rayhan	84,5	84	82	82,25	82,5	84	82,75	85,25	86,25	85,25	84,25	86,75	84,14583333	Sedang	
3	Aira Amelia	86,5	85	83	83,75	85	85	85,75	88,75	88	88	84,75	87,25	85,89583333	Sedang	
4	Bram Wibowo	84,75	86,5	87	82,75	84,75	87,75	86,5	84,75	85,75	85,5	85	86,75	85,64583333	Sedang	
5	Chirul Rifai	84,5	85	83	83,75	84,25	84,25	85	83,75	85,5	86,75	84,75	86,5	84,75	Sedang	
6	Iqbal Ramadhan	82,75	84,75	81,5	81,25	81,75	83	82,5	83,75	84,75	84,5	83,25	85,5	83,27083333	Sedang	
7	Kayla Aprilia Br. Siagian	85	86,25	85,5	82	80,75	85,5	87,25	86,75	85,75	87	84,5	89,5	85,47916667	Sedang	
8	M Reza Fahlevi Siregar	83,5	86,75	78,5	83,5	80	84	81,75	86,25	84,25	84	82,75	87,25	83,54166667	Sedang	
9	M Zidan Al Katiry Nasution	82,5	83,75	80	81,75	79	84	80,75	84,5	83,5	85,25	83	85,5	82,79166667	Sedang	
10	Marsha Sabina	84,5	83,25	79,5	80,5	80,25	83	83,5	85	85,25	85	83,75	86,5	83,33333333	Sedang	
11	Nurul Cahya Ramadhani	85	77,75	76,25	78	78	77,5	80,5	77,5	76,25	81,25	81	80,75	79,14583333	Rendah	
12	Puan Bunga Aurelie	89	85	86,5	84,75	87,75	88	89,25	88,5	88,5	88	88	87,25	87,54166667	Sedang	
13	Rasya Dwika Mailin Larosa	85,75	83,5	83	82,5	84,75	84,75	84,5	85,5	85,25	85,25	84,5	87,75	84,75	Sedang	
14	Rendi Ramadhan	83,5	83,25	79	81,5	80,5	83,75	82	84,75	84,5	85	83,75	86	83,125	Sedang	
15	Reza Aulia Waruwu	83,25	89	85,5	83	81,75	82,5	81,25	83,5	81,25	82,75	22,25	62,75	76,5625	Rendah	
16	Rifal Algi Fahri	81,25	81,5	78,75	78,75	79,25	82	80,5	78,25	81,5	81,25	62,5	65	77,54166667	Rendah	
17	Saka Syahridho	86,25	87,25	86,5	84	86,5	86,75	86,5	89	86,75	87,5	85,5	87,5	86,66666667	Sedang	
18	Satya Dwipi Farezi	82	84,5	79,25	81,25	79	82,75	82,25	83	84	84,25	82,5	85,75	82,54166667	Sedang	
19	Tio Ardiansyah	77,25	86	75	79	79,25	72	74	74	76,75	76	75,5	73,25	76,5	Rendah	
20	Uun Saqinah	86	86,5	85,5	83,75	87	88	87,25	88,5	87,5	88,75	88	87,25	87	Sedang	
21	Zais Revan Abimanyu	81,75	83,25	79,75	79,25	84,5	83,25	82,75	78,25	85,5	85	62,5	66,25	79,33333333	Rendah	
22	Aprilia Kartini	86	89	89,5	87	85,75	88,5	86,5	91,75	89,5	88,5	83,5	89,5	87,91666667	Sedang	
23	Cintiya Arista Putri	84	88	84,25	84,75	85,25	86	86	87	88	84,25	84,25	86,5	85,6875	Sedang	
24	Fitri Aulia Ramadani	82	82,25	75	79	76,75	75,5	74	71,25	76,75	78,25	74	79	76,97916667	Rendah	
25	Ian Ramadan Siahaan	82,75	86,75	81,5	79,5	81,25	83,5	83,75	85,75	86,75	87	82,75	85,75	83,91666667	Sedang	
26	Nabila Natasya	83,25	85,75	84,5	85,25	83,75	86	84,75	87,25	87,75	87,5	82,5	85,75	85,33333333	Sedang	
27	Nawan Dwi Okto	82,25	86,25	80,25	79,25	80,75	82,5	83	85,75	84,5	86,75	82,25	82,25	82,97916667	Sedang	
28	Radit Arya Pranata Saragih														Data tidak lengkap	
29	Rakha Aldiansyah	83,25	87,75	83	81,75	81,5	82	84	87,25	89,5	86,75	84,5	86,25	84,79166667	Sedang	
30	Rasya Akbar	83	86,75	82,75	80	82,75	84,25	85	87,25	88	88	82	88,25	84,83333333	Sedang	
31	Safira Mulhimah	88,25	81,75	77,75	78,75	78,75	74,75	77,25	76,5	77,5	76,5	77	77,5	78,52083333	Rendah	
32	Saskia Livia Ramadhani	85,75	89,5	88,75	85,25	86,25	87	87,75	92	89,75	90,25	85,75	87	87,91666667	Sedang	
33	Sri Rindiyani Hafizah	86,25	88,5	88	86,5	85,75	87,5	88,25	92	89,25	92	85	87	88	Sedang	
34	Susilo Darmawan	82,25	87,75	80,75	79,75	81	82	82,75	85,25	87,25	87,25	83	86	83,75	Sedang	
35	Syah Diza Aprilia	86,5	87,25	88,5	86,5	87,75	88,5	88,75	92,5	89,25	89,5	85,25	89	88,27083333	Sedang	
36	Abib Bin Razad	84,25	82,5	76,25	79,5	80	74,5	76	77,75	77,5	77,25	79	75,5	78,33333333	Rendah	
37	Rio Kadafi Nasution	85,5	88,25	85,5	80,75	84	85	85	86	87,5	87,25	85	86	85,47916667	Sedang	

	A	B	C
1	NAMA SISWA	RATA_RATA	KATEGORI
2	Syah Diza Aprillia	88,27083333	Sedang
3	Sri Rindiyani Hafizah	88	Sedang
4	Aprilia Kartini	87,91666667	Sedang
5	Saskia Livia Ramadha	87,91666667	Sedang
6	Puan Bunga Aurelie	87,54166667	Sedang
7	Uun Saqinah	87	Sedang
8	Saka Syahridho	86,66666667	Sedang
9	Aira Amelia	85,89583333	Sedang
10	Cintiya Arista Putri	85,6875	Sedang
11	Bram Wibowo	85,64583333	Sedang
12	Rio Kadafi Nasution	85,47916667	Sedang
13	Kayla Aprilia Br. Siagi	85,47916667	Sedang
14	Nabila Natasya	85,33333333	Sedang
15	Rasya Akbar	84,83333333	Sedang
16	Rakha Aldiansyah	84,79166667	Sedang
17	Rasya Dwika Mailin L	84,75	Sedang
18	Chirul Rifai	84,75	Sedang
19	Abdul Rayhan	84,14583333	Sedang
20	Ian Ramadan Siahaan	83,91666667	Sedang
21	Susilo Darmawan	83,75	Sedang
22	M Reza Fahlevi Sirega	83,54166667	Sedang
23	Marsha Sabina	83,33333333	Sedang
24	Iqbal Ramadhan	83,27083333	Sedang
25	Rendi Ramadhan	83,125	Sedang
26	Nawan Dwi Okto	82,97916667	Sedang
27	M Zidan Al Katiry Nas	82,79166667	Sedang
28	Satya Dwipi Farezi	82,54166667	Sedang
29	Zais Revan Abimanyu	79,33333333	Rendah
30	Nurul Cahya Ramadh	79,14583333	Rendah
31	Safira Mulhimah	78,52083333	Rendah
32	Abib Bin Razad	78,33333333	Rendah
33	Rifal Algi Fahri	77,54166667	Rendah
34	Fitri Aulia Ramadani	76,97916667	Rendah
35	Reza Aulia Waruwu	76,5625	Rendah
36	Tio Ardiansyah	76,5	Rendah
37	Radit Arya Pranata Saragih		Data tidak lengkap

Figure 1 Kmeans Agorithma Clustering Results

A. Results of the Tsukamoto Fuzzy Algorithm Implementation with the Program

The implementation of the Tsukamoto Fuzzy Algorithm was carried out by taking the ratios of each subject's scores from Semesters I to IV, then calculating the overall ratios to determine student achievement categories. Based on the calculated fuzzy agorithm, ratios below 60 are categorized as Low (weight 0), scores between 60 and 79 are categorized as Medium (weight 5), and scores above 80 are categorized as High (weight 10). The results of the data processing using the program indicate that most students are included in the High category with a weight of 10, while some students are included in the Middle category with a weight of 5, and only a few are classified as Low.

The implemented program successfully provided a more robust category change compared to K-Means, where the fuzzy naturaistic caculus directly assigned the vaues of the ratios to a certain weight. From the results, it was seen that students with ratios in the 80-80 range were consistently categorized as High, while students with ratios in the 60-79 range were included in the Medium category. With this approach, Fuzzy Logic provides a classification system that is easy to understand and in accordance with the predetermined fuzzy logic, making it easier for schools to group students' achievement levels.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	NAMA SISWA	PAI_AVG	PPKN_AVG	INDONESIA	HEMATIKA	IPA_AVG	IPS_AVG	INGGRIS_A1	BUDAYA	PJOK_AVG	AKARYA_A1	IQRA_AVG	TIK_AVG	RATA_RATA	KATEGORI	BOBOT	
2	Abdul Rayhan	84,5	84	82	82,25	82,5	84	82,75	85,25	86,25	85,25	84,25	86,75	84,14583333	Tinggi	10	
3	Aira Amelia	86,5	85	83	83,75	85	85	85,75	88,75	88	88	84,75	87,25	85,89583333	Tinggi	10	
4	Bram Wibowo	84,75	86,5	87	82,75	84,75	87,75	86,5	84,75	85,75	85,5	85	86,75	85,64583333	Tinggi	10	
5	Chirul Rifal	84,5	85	83	83,75	84,25	84,25	85	83,75	85,5	86,75	84,75	86,5	84,75	Tinggi	10	
6	Iqbal Ramadhan	82,75	84,75	81,5	81,25	81,75	83	82,5	83,75	84,75	84,5	83,25	85,5	83,27083333	Tinggi	10	
7	Kayla Aprilia Br. Siagian	85	86,25	85,5	82	80,75	85,5	87,25	86,75	85,75	87	84,5	89,5	85,47916667	Tinggi	10	
8	M Reza Fahlevi Siregar	83,5	86,75	78,5	83,5	80	84	81,75	86,25	84,25	84	82,75	87,25	83,54166667	Tinggi	10	
9	M Zidan Al Katiry Nasution	82,5	83,75	80	81,75	79	84	80,75	84,5	83,5	85,25	83	85,5	82,79166667	Tinggi	10	
10	Marsha Sabina	84,5	83,25	79,5	80,5	80,25	83	83,5	85	85,25	85	83,75	86,5	83,33333333	Tinggi	10	
11	Nurul Cahya Ramadhani	85	77,75	76,25	78	78	77,5	80,5	77,5	76,25	81,25	81	80,75	79,14583333	Tinggi	10	
12	Puan Bunga Aurelie	89	85	86,5	84,75	87,75	88	89,25	88,5	88,5	88	88	87,25	87,54166667	Tinggi	10	
13	Rasya Dwika Mailin Larosa	85,75	83,5	83	82,5	84,75	84,75	84,5	85,5	85,25	85,25	84,5	87,75	84,75	Tinggi	10	
14	Rendi Ramadhan	83,5	83,25	79	81,5	80,5	83,75	82	84,75	84,5	85	83,75	86	83,125	Tinggi	10	
15	Reza Aulia Waruwu	83,25	89	85,5	83	81,75	82,5	81,25	83,5	81,25	82,75	22,25	62,75	76,5625	Sedang	5	
16	Rifal Algi Fahr	81,25	81,5	78,75	78,75	79,25	82	80,5	78,25	81,5	81,25	62,5	65	77,54166667	Sedang	5	
17	Saka Syahritho	86,25	87,25	86,5	84	86,5	86,75	86,5	89	86,75	87,5	85,5	87,5	86,66666667	Tinggi	10	
18	Satya Dwipi Farezi	82	84,5	79,25	81,25	79	82,75	82,25	83	84	84,25	82,5	85,75	82,54166667	Tinggi	10	
19	Tio Ardiansyah	77,25	86	75	79	79,25	72	74	74	76,75	76	75,5	73,25	76,5	Sedang	5	
20	Uun Saqinah	86	86,5	85,5	83,75	87	88	87,25	88,5	87,5	88,75	88	87,25	87	Tinggi	10	
21	Zais Revan Abimanyu	81,75	83,25	79,75	79,25	84,5	83,25	82,75	78,25	85,5	85	62,5	66,25	79,33333333	Tinggi	10	
22	Aprilia Kartini	86	89	89,5	87	85,75	88,5	86,5	91,75	89,5	88,5	83,5	89,5	87,91666667	Tinggi	10	
23	Cintiya Arista Putri	84	88	84,25	84,75	85,25	86	86	87	88	84,25	84,25	86,5	85,6875	Tinggi	10	
24	Fitri Aulia Ramadani	82	82,25	75	79	76,75	75,5	74	71,25	76,75	78,25	74	79	76,97916667	Sedang	5	
25	Ian Ramadan Siahaan	82,75	86,75	81,5	79,5	81,25	83,5	83,75	85,75	86,75	87	82,75	85,75	83,91666667	Tinggi	10	
26	Nabila Natasya	83,25	85,75	84,5	85,25	83,75	86	84,75	87,25	87,5	87,5	82,5	85,75	85,33333333	Tinggi	10	
27	Nawan Dwi Okto	82,25	86,25	80,25	79,25	80,75	82,5	83	85,75	84,5	86,75	82,25	82,25	82,97916667	Tinggi	10	
28	Radit Arya Pranata Saragih														Data tidak lengkap		
29	Rakha Aldiansyah	83,25	87,75	83	81,75	81,5	82	84	87,25	89,5	86,75	84,5	86,25	84,79166667	Tinggi	10	
30	Rasya Akbar	83	86,75	82,75	80	82,75	84,25	85	87,25	88	88	82	88,25	84,83333333	Tinggi	10	
31	Safira Mulhimah	88,25	81,75	77,75	78,75	78,75	74,75	77,25	76,5	77,5	76,5	77	77,5	78,52083333	Sedang	5	
32	Saskia Livia Ramadhani	85,75	89,5	88,75	85,25	86,25	87	87,75	92	89,75	90,25	85,75	87	87,91666667	Tinggi	10	
33	Sri Rindiyani Hafizah	86,25	88,5	88	86,5	85,75	87,5	88,25	92	89,25	92	85	87	88	Tinggi	10	
34	Susilo Darmawan	82,25	87,75	80,75	79,75	81	82	82,75	85,25	87,25	87,25	83	86	83,75	Tinggi	10	
35	Syah Diza Aprililia	86,5	87,25	88,5	86,5	87,75	88,5	88,75	92,5	89,25	89,5	85,25	89	88,27083333	Tinggi	10	
36	Abib Bin Razad	84,25	82,5	76,25	79,5	80	74,5	76	77,75	77,5	77,25	79	75,5	78,33333333	Sedang	5	
37	Rio Kadafi Nasution	85,5	88,25	85,5	80,75	84	85	85	86	87,5	87,25	85	86	85,47916667	Tinggi	10	

	A	B	C	D	E
1	NAMA SISWA	RATA_RATA	KATEGORI	BOBOT	
2	Syah Diza Aprillia	88,27083333	Tinggi	10	
3	Sri Rindiyan Hafizah	88	Tinggi	10	
4	Aprilia Kartini	87,91666667	Tinggi	10	
5	Saskia Livia Ramadhani	87,91666667	Tinggi	10	
6	Puan Bunga Aurelie	87,54166667	Tinggi	10	
7	Uun Saqinah	87	Tinggi	10	
8	Saka Syahridho	86,66666667	Tinggi	10	
9	Aira Amelia	85,89583333	Tinggi	10	
10	Cintiya Arista Putri	85,6875	Tinggi	10	
11	Bram Wibowo	85,64583333	Tinggi	10	
12	Rio Kadafi Nasution	85,47916667	Tinggi	10	
13	Kayla Aprilia Br. Siagian	85,47916667	Tinggi	10	
14	Nabila Natasya	85,33333333	Tinggi	10	
15	Rasya Akbar	84,83333333	Tinggi	10	
16	Rakha Aldiansyah	84,79166667	Tinggi	10	
17	Rasya Dwika Mailin Larosa	84,75	Tinggi	10	
18	Chirul Rifai	84,75	Tinggi	10	
19	Abdul Rayhan	84,14583333	Tinggi	10	
20	Ian Ramadan Siahaan	83,91666667	Tinggi	10	
21	Susilo Darmawan	83,75	Tinggi	10	
22	M Reza Fahlevi Siregar	83,54166667	Tinggi	10	
23	Marsha Sabina	83,33333333	Tinggi	10	
24	Iqbal Ramadhan	83,27083333	Tinggi	10	
25	Rendi Ramadhan	83,125	Tinggi	10	
26	Nawan Dwi Okto	82,97916667	Tinggi	10	
27	M Zidan Al Katiry Nasution	82,79166667	Tinggi	10	
28	Satya Dwipi Farezi	82,54166667	Tinggi	10	
29	Zais Revan Abimanyu	79,33333333	Tinggi	10	
30	Nurul Cahya Ramadhani	79,14583333	Tinggi	10	
31	Safira Mulhimah	78,52083333	Sedang	5	
32	Abib Bin Razad	78,33333333	Sedang	5	
33	Rifal Algi Fahri	77,54166667	Sedang	5	
34	Fitri Aulia Ramadani	76,97916667	Sedang	5	
35	Reza Aulia Waruwu	76,5625	Sedang	5	
36	Tio Ardiansyah	76,5	Sedang	5	
37	Radit Arya Pranata Saragih		Data tidak lengkap		
--					

Figure 2 Results of Algorithm Grouping Fuzzy Tsukamoto

B. Discussion of Comparison of K-Means and Fuzzy Tsukamoto Results

The comparison of the K-Means algorithm with Fuzzy Logic shows the fundamental differences in the concept of the output that is being processed. The K-Means algorithm works with the principle of unsupervised clustering based on the Euclidean path, so that the position of the centroid corresponds to the distribution of the students. With centroid parameters of 75, 85, and 95, the clustering divides the students into three groups (Low, Middle, High) more proportionally. In contrast, the Fuzzy Tsukamoto algorithm uses a simple linguistic algorithm based on the weight of the values, namely <60 as Low (weight 0), 60–79 as Medium (weight 5), and ≥ 80 as High (weight 10). With this algorithm, the categorization becomes more precise, as the categorization is not affected by the distribution of values, but is entirely determined by the weight of the values applied.

In terms of the distribution of the result, K-Means is more balanced because students with ratios around 84–86 in the Middle category and more close to the centroid 85. Meanwhile, Fuzzy Tsukamoto directly categorizes all students with ratios ≥ 80 into the High category. This can be seen in the example of three students, namely Abdul Rayhan (84.1), Aira Amelia (85.9), and Chirul Rifai (84.8), which by K-Means are categorized into the cluster of result position near the centroid 85, while by Fuzzy Tsukamoto are categorized as high result value passing through the limit ≥ 80 . Thus, K-Means is more flexible in grouping the distribution based on the distribution, while Fuzzy Tsukamoto provides a simpler classification in accordance with the specified nature.

Concept Comparison

Aspek	K-Means	Fuzzy Tsukamoto
Theoretica Basis	Unsupervised clustering based on Euclidean distance.	Simple fuzzy inference based on linguistic rules.
Parameters	Initial centroid (75, 85, 95). Can "adjust" the centroid to the actual data;	Cutoff rule (60, 79, 80). The resulting category is fixed according to the rule; all scores ≥ 80 are automatically High.
Cluster Results	for example, the Medium cluster averages around 84–85. More balanced (some students are in the Medium cluster even though their scores are 84–85, because they are closer to the 85 centroid).	Almost all students are categorized as "High" because the average score in the dataset is above 80.
Student Distribution		
Data Sensitivity	Flexible: if many students score 70–75, the lower centroid will shift.	Rigid: Fixed cutoffs of 60–79 & ≥ 80 despite the imbalanced data distribution.
Output	Cluster ID + category + final centroid.	Linguistic categories + discrete weights (0, 5, 10)

Perbandingan Hasil (3 siswa)

Name	Average	K-Means (cluster)	Fuzzy Tsukamoto
Abdul Rayhan	84.1	Moderate (C ₂ near 85)	Tal ($\geq 80 \rightarrow 10$)
Aira Amelia	85.9	Moderate (C ₂ ~85)	Tal ($\geq 80 \rightarrow 10$)
Chirul Rifai	84.8	Moderate (C ₂ ~85)	Tal ($\geq 80 \rightarrow 10$)

K-Methods can classify students with scores around 84–86 as Middle, while the centroid for High (95) is too far. Fuzzy Tsukamoto immediately categorizes all scores ≥ 80 as High, even though it takes into account the relative distribution of students. So the conclusion is that if the goal is to group relatively close friends within the same class, K-Methods is more suitable, it can distinguish "Middle vs High" even though all scores are high. Jika tujuanmu klasifikasi mutlak berdasar standar nilai Fuzzy Tsukamoto lebih cocok, karena aturan sudah jelas (cut-off ≥ 80 = Tinggi).

Discussion

The comparative analysis between the K-Means and Fuzzy Tsukamoto algorithms in clustering student academic achievement at SMP NU Medan reveals significant methodological and conceptual differences that influence the interpretation and applicability of the results in educational management. The implementation of the K-Means algorithm produced three main clusters—Low, Medium, and High—based on centroid positions at 75, 85, and 95, respectively. The calculation process used the average ratio of each student's scores from semesters I to IV, resulting in a distribution where most students fell into the Medium category with values centered around the centroid of 85. This indicates that K-Means effectively represents the relative performance distribution among students, offering a flexible clustering structure that adjusts according to data density. In contrast, the Fuzzy Tsukamoto algorithm categorized students into three linguistic groups—Low (≤ 60), Medium (61–79), and High (≥ 80)—based on fuzzy membership rules and assigned corresponding weights of 0, 5, and 10. The results showed that most students were classified into the High category due to the overall high score distribution, reflecting the deterministic nature of the fuzzy-based rule system. While K-Means demonstrates adaptability to data variations and provides a more balanced representation of clusters, Fuzzy Tsukamoto offers greater interpretability and simplicity by establishing clear, rule-based thresholds. From a decision-making perspective, K-Means is more appropriate for identifying relative performance differences within a cohort, enabling teachers to design differentiated instructional strategies for groups that fall near specific centroids. Meanwhile, Fuzzy Tsukamoto is more suitable for absolute classification based on predetermined performance standards, making it useful for administrative evaluations such as determining eligibility for awards or interventions. Thus,

both algorithms have complementary strengths: K-Means excels in detecting data-driven group patterns, while Fuzzy Tsukamoto enhances transparency and ease of interpretation. The combination of these two methods can provide a more comprehensive analysis framework that not only identifies performance clusters but also translates them into actionable insights for educational decision-making, aligning with the school's goal of improving teaching quality and optimizing student development strategies through data-based approaches.

4. Conclusion

Based on the comprehensive analysis, calculation, and program implementation using the K-Means and Fuzzy Tsukamoto algorithms in clustering student academic achievement at SMP NU Medan, it can be concluded that both approaches offer complementary strengths in educational data analysis and decision-making. The K-Means algorithm effectively grouped students into three achievement categories-Low, Medium, and High-based on predetermined centroids (75, 85, 95), demonstrating its capability to identify relative performance clusters with proportional and stable centroid distributions. This adaptive characteristic enables more precise differentiation among students, particularly those whose average scores fall near cluster boundaries, providing valuable insights for individualized learning strategies. Conversely, the Fuzzy Tsukamoto algorithm, utilizing simple fuzzy weights (0 for Low, 5 for Medium, and 10 for High), offers a rule-based classification system that emphasizes interpretability and aligns with fixed academic performance standards. Empirical testing revealed that most students with average scores above 80 were automatically classified as "High" with a fuzzy weight of 10, reflecting the algorithm's deterministic and threshold-based nature. The comparative findings suggest that K-Means is more suitable for assessing relative performance within heterogeneous student populations, while Fuzzy Tsukamoto is more effective for absolute performance evaluation and administrative assessments. In practical application, both algorithms can be efficiently implemented using accessible tools such as Excel and Python, facilitating their integration into school data management systems. The implications of this study underscore the potential of combining data-driven clustering and fuzzy logic approaches to enhance educational analytics, support data-informed instructional planning, and improve student performance evaluation. However, this study's limitation lies in its use of data from a single institution, which may affect the generalizability of results. Future research should explore hybrid or ensemble-based clustering models across multiple schools and diverse datasets to strengthen robustness, improve predictive accuracy, and develop adaptive frameworks that can better capture dynamic learning patterns in broader educational contexts.

References

- Adawiyah, Q., Defit, S., & Sumijan. (2024). Penerapan Algoritma K-Means Clustering untuk Mengelompokkan Rekomendasi Metode Kontrasepsi Berbasis Machine Learning di Puskesmas. *Jurnal KomtekInfo*, 11(4), 300–305. <https://doi.org/10.35134/komtekinfo.v11i4.563>
- Adzika, A. K., & Djagba, P. (2024). Inference with K-means. <http://arxiv.org/abs/2410.17256>
- Anastassia, S., Kharis, A., Haqqi, A., & Zili, A. (2024). Prosiding Seminar Nasional Sains dan Teknologi Seri 02 Fakultas Sains dan Teknologi. Universitas Terbuka, 1(2).
- Ardianti, M., Nurhayati, O. D., & Warsito, B. (2024). Model Prediksi Kinerja Siswa Berdasarkan Data Log LMS Menggunakan Ensemble Machine Learning. *JST (Jurnal Sains dan Teknologi)*, 12(3). <https://doi.org/10.23887/jstundiksha.v12i3.59816>
- Ariesanto Akhmad, E. P. (2020). Data Mining Menggunakan Regresi Linear untuk Prediksi Harga Saham Perusahaan Pelayaran. *Jurnal Aplikasi Pelayaran dan Kepelabuhanan*, 10(2), 120. <https://doi.org/10.30649/japk.v10i2.83>
- Ayu, D., Wulandari, N., & Prasetyo, A. (2018). Sistem Penunjang Keputusan Untuk Menentukan Status Gizi Balita Menggunakan Metode Fuzzy Tsukamoto. *Jurnal Informatika*, 5(1), 22–33.
- Eka Agustin, F. M., Fitria, A., & Hanifah, A. S. (2015). Implementasi Algoritma K-Means untuk Menentukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional (Studi Kasus: SMP Negeri 101 Jakarta). 8(1).
- Elisna, F., Zunaidi, M., & Erwansyah, K. (n.d.). Penerapan Data Mining Untuk Menganalisa Tingkat Kepuasan Pelanggan Telkomsel Terhadap Sikap Pelayanan Caroline Officer Dengan Metode Algoritma K-Means Clustering. *Jurnal CyberTech*.

- Fernando Ade Pratama, E., & Jumadi, J. (n.d.). Kampus I: Jl Meranti Raya No.32 Sawah Lebar Kota Bengkulu 38228 Telp. (0736) 22027, Fax. Jurnal Media Infotama, 18(2), 341139.
- Gupta, N. S., Agrawal, B., Chauhan, R. M., & Mehta, A. B. (2015). Survey on Clustering Techniques of Data Mining. American International Journal of Research in Science, 15-188. <http://www.iasir.net>
- Karlina, L., & Nurdiawan, O. (2023). Penerapan K-Medoids dalam Klasifikasi Persebaran Lahan Kritis di Jawa Barat Berdasarkan Kabupaten/Kota. Jurnal Mahasiswa Teknik Informatika, 7(1). <https://opendata.jabarprov.go.id/id>
- Liu, Y., Ma, S., & Du, X. (2024). A Novel Effective Distance Measure and a Relevant Algorithm for Optimizing the Initial Cluster Centroids of K-means. IEEE Access. <https://doi.org/10.1109/ACCESS.2020.3044069>
- Marcelina, D., Kurnia, A., & Terttiaavini, T. (2023). Analisis Kluster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 3(2), 293-301. <https://doi.org/10.57152/malcom.v3i2.952>
- Nasution, Y. R., & Eka, M. (2018). Penerapan Algoritma K-Means Clustering pada Aplikasi Menentukan Berat Badan Ideal. Algoritma: Jurnal Ilmu Komputer dan Informatika.
- Nur, F., Fauzan, R., Aziz, J., Darma Setiawan, B., & Arwani, I. (2018). Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa. 2(6). <http://j-ptiik.ub.ac.id>
- Nurul Fatma Dewi Mardianto, & Yahfizham, Y. (2024). Systematic Literature Review: Penerapan Berpikir Komputasi Dalam Pembelajaran Matematika. Journal of Student Research, 2(4), 41-55. <https://doi.org/10.55606/jsr.v2i4.3082>
- Oktario Dacwanda, D., & Nataliani, Y. (2021). Implementasi K-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan. AITI: Jurnal Teknologi Informasi, 18(2), 125-138.
- Papakyriakou, D., & Barbounakis, I. S. (2022). Data Mining Methods: A Review. International Journal of Computer Applications, 183(48).
- Priya, P., & Souza, D. A. D. (n.d.). Analysis of K-Means Clustering Based Image Segmentation. 10(2), 1-6. <https://doi.org/10.9790/2834-10210106>
- Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. JEPIN (Jurnal Edukasi dan Penelitian Informatika).
- Pujjarso Nugroho, R., Darma Setiawan, B., & Tanzil Furqon, M. (2019). Penerapan Metode Fuzzy Tsukamoto untuk Menentukan Harga Sewa Hotel (Studi Kasus: Gili Amor Boutique Resort, Dusun Gili Trawangan, Nusa Tenggara Barat). 3(3). <http://j-ptiik.ub.ac.id>
- Putra, E. M. I., & Syafrullah, M. (2018). Implementasi Algoritma ST-DBSCAN dan K-Means untuk Pengelompokan Indeks Pembangunan Manusia Kabupaten/Kota Pulau Jawa Tahun 2014-2016 Berbasis Web di Badan Pusat Statistik. 1(3).
- Putra, R. R., & Wadisman, C. (2018). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means. Journal of Information Technology and Computer Science (INTECOMS), 1(1).
- Putri, D., Asih, A., Irawan, B., & Bahtiar, A. (2024). Pengelompokan Data Transaksi dalam Menentukan Strategi Penjualan Menggunakan Algoritma K-Means. Jurnal Mahasiswa Teknik Informatika, 8(1).
- Rezky, M. P., Sutarto, J., Prihatin, T., Yulianto, A., Haidar, I., & Surel, A. (2019). Prosiding Seminar Nasional Pascasarjana UNNES.
- Rizal, S. (2019). Development of Big Data Analytics Model. ITEJ, 4(1).
- Saligkaras, D., & Papageorgiou, V. E. (n.d.). Seeking the Truth Beyond the Data: An Unsupervised Machine Learning Approach.
- Sapriadi, S., Hayati, N., Syaputra, A. E., Eirlangga, Y. S., Manurung, K. H., & Hayati, N. (2023). Sistem Pakar Diagnosa Gaya Belajar Mahasiswa Menggunakan Metode Forward Chaining. Jurnal Informasi dan Teknologi, 5(3), 71-78. <https://doi.org/10.60083/jidt.v5i3.381>

- Yuli Mardi. (n.d.). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. Jurnal Edik Informatika.
- Yunita, F. (2018). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering pada Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Islam Indragiri). Jurnal Sistemasi, 7(1).
- Zuhal, N. K. (2022). Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering. Universitas Nusantara PGRI Kediri, 1(1).