Fuzzy Clustering-Based Grouping for Mapping the Distribution of Student Success Data

Mustakim[1*], Delvi Nur Aini[2],
Ana Uzla Batubara[3], Moh. Erkamim[4], Legito[5]

[1,2] Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
[3]Sekolah Tinggi Ilmu Kesehatan Binalita Sudama Medan, Indonesia
[4]Program Studi Sistem Informasi Kota Cerdas, Universitas Tunas Pembangunan Surakarta, Indonesia
[5]Jurusan Teknik Informatika, Sekolah Tinggi Teknologi Sinar Husni Deliserdang, Sumatera Utara, Indonesia

Email: [1]mustakim@uin-suska.ac.id, [2]12050320493@students.uin-suska.ac.id,
[3]anauzla@gmail.com, [4]erkamim@lecture.utp.ac.id, [5]legitostt@gmail.com

Abstract

Learning activities are the main activity in the overall teaching and learning process in schools. This is because whether or not the achievement of educational goals depends on how the learning process is carried out by students. The uneven level of student success in learning is one of the problems in the school's efforts to realize the vision and mission of SMKN 5 Pekanbaru in preparing skilled graduates to be able to work in certain sectors by the public interest and the industrial world. In this study, mapping and grouping student grade data was carried out using the Fuzzy C-Means algorithm to provide information to the school in making the right decisions and optimizing the learning process. Furthermore, clustering was carried out in several experiments K=3 to K=7, and obtained the best validity value tested with the Silhouette Index of 0.4277 located at K=5. Then the distribution of cluster 5 on student score data was obtained with details, namely cluster 1 with a capacity of 1 student, cluster 2 with a capacity of 27 students, cluster 3 with a capacity of 1 student, cluster 4 with a capacity of 10 students, cluster 5 with a capacity of 23 students.

Keyword: Clustering, Fuzzy C-Means, Silhouette Index, The Success of Students

1. INTRODUCTION

Vocational High School is one manifestation of formal education that aims to equip graduates with knowledge and skills beneficial for themselves in their professions and public interests. Therefore, there is a need for efforts to improve the quality of graduates, including fostering good collaboration and coordination between the government, schools, cooperation in the industrial sector, teacher expertise, the role of parents, and the students themselves [1][2].

Based on data from the Education Department of Riau Province (2022), Vocational High School or SMKN 5 Pekanbaru is a secondary school located on Yos Sudarso Street, Rumbai District, Pekanbaru City, Riau. The school was established in 1995 and currently offers 12 study programs. SMKN 5 Pekanbaru is one of the vocational schools registered in the "Center of Excellence Vocational School" program, which includes several programs such as strengthening human resources (HR) in both Hard Skills and Soft Skills aspects. The program prioritizes student achievements, graduate skills, and student success in learning [3].

The uneven level of student success in learning becomes one of the challenges in the school's efforts to realize the vision and mission of SMKN 5 Pekanbaru in preparing skilled graduates for a career in specific sectors according to the needs of the community and the industrial field [2]. To gain a better understanding of the factors that have the greatest impact on student success, schools require a technique or method that can help them analyze data and make informed decisions. Data mining is an effective method that can be used for this purpose [4][1].

Data mining is a process that applies mathematical, statistical, artificial intelligence, and machine learning methods to extract useful facts from a dataset [5][6]. To facilitate the grouping of the distribution of student success data, a technique for grouping the data (clustering) is required. The Fuzzy Clustering algorithm or Fuzzy C-Means (FCM) is one type of algorithm applied in clustering data, where the degree of membership determines the presence of each data point in a group [7].

Extensive research has been conducted on clustering techniques for grouping data. The Fuzzy C-Means algorithm has the highest silhouette index accuracy comparison of 0.2559, which is better than the k-medoids algorithm [8] and has an accuracy rate of 76% when compared to the k-means algorithm [7]. The Fuzzy C-Means algorithm is excellent at detailed grouping [9] based on the data attributes it possesses [10], and it performs exceptionally well in finding high-level clusters [11] Additionally, it can map interactions between different cluster shapes [12] and demonstrate test results with a best percentage of 71.23% [13]. The Silhouette Index is shown to have a strong structure category value of 0.84321191, indicating that the Fuzzy C-Means algorithm is well-suited for grouping product data into three clusters [14].

Another study in grouping student achievement using the Fuzzy C-Means Clustering algorithm resulted in a validity value of 1.3554 which mapped as many as 3 clusters based on academic scores [15]. The Fuzzy C-Means method also proved optimal in cluster center placement compared to other cluster methods in predicting student truancy factors [16].

Based on the problem statement and referring to previous research, data mining techniques using the Fuzzy Clustering algorithm are applied for mapping the distribution of student success data at SMKN 5 Pekanbaru. This study also aims to assess the ability and reliability of the Fuzzy Clustering algorithm, with the hope of assisting the school in making informed decisions and optimizing the learning process at SMKN 5 Pekanbaru.

2. MATERIALS AND METHOD
2.1. Data

Data collection in this study includes a dataset of report card scores of grade 12 students in the Department of Building Modeling and Information Design (DPIB) for the 2019-2020 Academic Year, sourced from the Curriculum Department Database at SMKN 5 Pekanbaru. . The data was collected from July 15 to August 31, 2022, with a total of 62 data records with details in Table 1.

Table 1 . Student Grade Data from *Cleaning*

Name	Gender	Religion	Religion Edu.	Pancasila Edu.	...	EBK	KUG	KJJ
Student 1	1	1	76	80	...	78	78	73
Student 2	2	1	77	83	...	83	87	80
Student 3	1	1	76	76	...	75	76	71
Student 4	2	1	80	84	...	95	89	83
Student 5	1	1	76	81	...	80	81	71
Student 6	1	1	78	82	...	80	84	72
Student 7	1	1	80	83	...	85	90	80
...	...	...	...	...	...	...	...	...
Students 62	2	1	82	81	...	83	90	78

2.2. Method

The method applied in this study consists of several stages: Data Collection, Data Pre-Processing, Clustering, Validity Comparison, Results, and Analysis. The research begins with the process of data accumulation (preprocessing) carried out to clean data or delete data that is missing value. The results of preprocessing data are then normalized which aims to produce a balance between low and high values in finding the initial value, maximum, and minimum data. It then applies the Fuzzy C-Means algorithm to map student grade data and generate clusters of data to be evaluated using the Silhouette Index Validity. The process and results of the analysis are carried out to compare the validity values obtained in the final analysis clustering. Research Methodology described in Figure 1.

2.3. Data Mining

Data mining is a method in which one or more computer system learning techniques are instructed to acquire and analyze knowledge in a series of stages to add value to a database that is not obtained manually [17]. Data mining is a technique for discovering valuable facts in a sequential set of databases to detect previously unknown trends, patterns, and relationships that can be utilized to form predictive models [18][19] and aims to detect specific models that need to be obtained for use in valid decision-making in the future[20].

2.4. Clustering

lustering is a process of categorizing data based on their level of similarity. This process aims to maximize the similarity among objects within a cluster while minimizing the similarity between different clusters. The similarity between two objects is measured by the relationship between them [21]. Two types of data clustering techniques are commonly used: hierarchical clustering and non-hierarchical clustering.

Hierarchical clustering involves classifying objects to obtain the total number of clusters required. On the other hand, non-hierarchical clustering begins by specifying the desired number of clusters. Fuzzy clustering is a widely used non-hierarchical clustering technique [22].
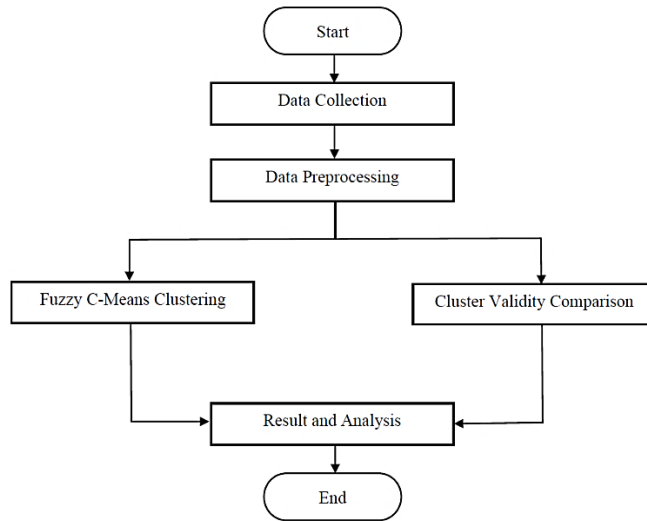

Figure 1. Research Methodology

2.5. Fuzzy C-Means

Fuzzy C-Means (FCM) is one clustering method that considers the presence of data in a cluster determined by the position of members that encompass a fuzzy set. This can help researchers determine the similarity of each data object in a group. Fuzzy C-Means fall under unsupervised learning methods with a predetermined total cluster in the clustering process. The algorithm commonly used in Fuzzy Clustering is FCM [23].

The calculation method in the Fuzzy C-Means algorithm is as follows:
1. Cluster X input data, which consists of a matrix with dimensions n x m (n, total sample data; m, attributes of each data). Xij, i-th sample data (i=1,2...n), j-th attribute (j=1,2...m).

2. Specify:
    a. Total Cluster: c;
    b. Rank: w;
    c. Maximum Iteration: MaxIter;
    d. Error Rate: εI;
    e. Initial Objective Function: P o= 0;
    f. Initial Iteration: t = 1;

3. Decrease random numerical μik, i=1,2...n; k=1.2...c; into the elements of the initial partition matrix U.

$$Qi \ = \ \Sigma = c\ k\ \mu ik \tag{1}$$

With j = 1, 2, ... n
Count:

$$\mu ik \ = \ \mu ik\ Qi \tag{2}$$

4. Specify the k-th cluster center point: Vkj, where k=1.2...c; and j=1.2...m

$$V_{kj} \ = \ \frac{\sum_{i}^{n}=1\ ((\mu ik)w_* X_{ij}}{\sum_{i}^{n}=1\ (\mu ik)^{w}} \tag{3}$$

5. Define objective functions according to the t-th iteration, Pt

$$Pt \ = \ \sum_{i}^{n}=1 \sum_{k}^{c}=1 \left(\left[\sum_{j}^{m}=1 (X_{ij}-V_{ij})^2 \right] (\mu ik)^2 \right) \tag{4}$$

6. Calculate transformations on partition matrix

$$\mu ik \; = \; \frac{\left[\sum_{j=1}^{m}\left(X_{ij}-V_{kj}\right)^2\right]^{\frac{-1}{w-1}}}{\left[\sum_{k=}^{c}\left(X_{ij}-V_{kj}\right)^2\right]^{\frac{-1}{w-1}}} \tag{5}$$

7. Check stop position:
   If: ($|Pt\text{-}Pt\text{-}1| < \varepsilon_I$) or (t > MaxIter) then stop. If not: t = t + 1, repeat at the 4th step

2.6. Silhouette Index (SI)

The analysis of the SI method can be applied for validation in the Fuzzy C-Means algorithm. SI is used to estimate the average distance values between clusters in assessing the quality of the clusters and predicting how well a study becomes a single cluster. This method is performed by estimating the similarity or dissimilarity between shapes within one cluster and another cluster from the analysis results [24][25]. A cluster is considered more optimal if it has a value that approaches 1[12]. The following is the formula for calculating the SI value in equation 6.

$$SI \; = \; \frac{1}{n}\sum_{i=1}^{n}\left(\frac{b(i)-a(i)}{max\ \{a(i),b(i)\}}\right) \tag{6}$$

2.7. The Success of Students

Learning activities are a fundamental aspect of the overall teaching and learning process in schools. This is because the achievement of educational goals depends on how students engage in the learning process. To realize students' success in the teaching and learning activities, the role of the teacher is crucial, acting as both a motivator and a facilitator. Implementing effective and understandable teaching methods is one type of effort that can be made by teachers. Furthermore, the adopted teaching methods also influence the success of the teaching and learning process in achieving the intended goals [26][27].

The role of a teacher in the process of learning is quite complex. They are not just responsible for delivering course materials, but also for encouraging and guiding students to become more enthusiastic about learning. In order to achieve this, teachers must be proficient in the subject matter they are teaching and use effective delivery models. Additionally, they should continuously strive to develop their skills and create a conducive learning environment to ensure the success of their students in their learning endeavors [22] [24].

2.8. Tools

This study uses tools, namely RapidMiner for the cluster formation process. RapidMiner is an application with a GUI (Graphical User Interface) that makes it easier for users to use this software. This software is open source and created using Java programs under the GNU Public License [29]. This study also uses Matlab tools for cluster validation process using silhouette index validity. MATLAB (Matrix Laboratory) is a high-level programming language specifically used for numerical computing, programming, and visualization. Matlab allows matrix manipulation, plotting functions and data, implementing algorithms, creating user interfaces, and interfaceing with programs in other languages [30]

3. RESULTS AND DISCUSSION
3.1. Preprocessing Data Result

The results of preprocessing the data resulted in the attributes used covering gender, religion, and a total of 11 subjects. The results of the data cleaning process are then carried out in the normalization process. The normalization applied is the min-max normalization method. The following are the results of the data normalization process presented in Table 2.

Table 2. Normalization of Student Grade Data

Name	Gender	Religion	Religion Edu.	Pancasila Edu.	...	EBK	KUG	KJJ
Student 1	0.0000	0.0000	0.5128	0.4615	...	0.1500	0.1765	0.2308
Student 2	1.0000	0.0000	0.5385	0.6923	...	0.4000	0.7059	0.7692
Student 3	0.0000	0.0000	0.5128	0.1538	...	0.0000	0.0588	0.0769
Student 4	1.0000	0.0000	0.6154	0.7692	...	1.0000	0.8235	1.0000
Student 5	0.0000	0.0000	0.5128	0.5385	...	0.2500	0.3529	0.0769
Student 6	0.0000	0.0000	0.5641	0.6154	...	0.2500	0.5294	0.1538
Student 7	0.0000	0.0000	0.6154	0.6923	...	0.5000	0.8824	0.7692
...	...	...	...	...	...	...	...	...
Students 62	1.0000	0.0000	0.6667	0.5385	...	0.4000	0.8824	0.6154

3.2. Clustering with Fuzzy C-Means Algorithm

In this study, the Fuzzy C-Means algorithm was used to cluster data that had undergone pre-processing and normalization in the previous stage. The objective of the clustering tests was to obtain the most appropriate cluster mapping. The Fuzzy C-Means algorithm was tested with different parameters, including the number of clusters (ranging from 3 to 7) and the power input set to 2, with Maxiter set to 100. The results of the clustering mapping are presented in Table 3.

Table 3. Fuzzy C-Means clustering results

Experiment	Cluster						
	1	2	3	4	5	6	7
K=3	11	23	28	-	-	-	-
K=4	0	23	23	16	-	-	-
K=5	1	27	1	10	23	-	-
K=6	4	26	0	24	8	0	-
K=7	12	6	11	10	8	5	10

3.3. Cluster Validity Comparison

Based on clustering experiments with Fuzzy C-Means algorithm on experiments 3 clusters into 7 clusters. The best cluster results are obtained by comparing the validity values obtained using SI in each cluster experiment. The following comparative results of the experiments that have been carried out are presented in Figure 2.
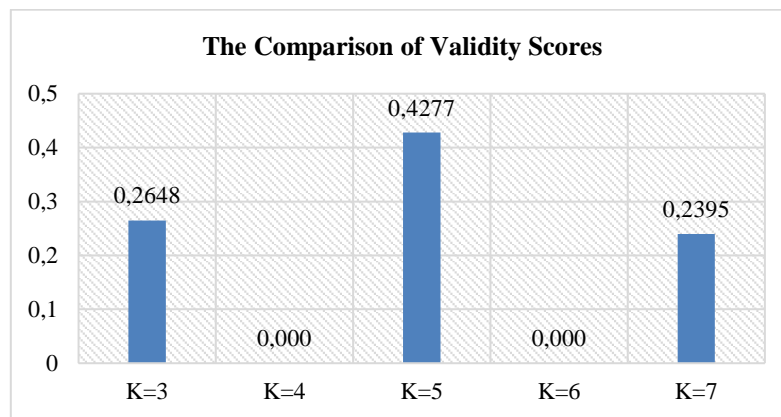

Figure 2. Validity Value Comparison

The test value results are presented in Figure 2. The validity of the entire cluster model experiment lies in the range of K=3 to K=7. Among these, the validity value at K=5 is the most optimal for cluster placement, as it is closest to 1. Therefore, the SI validity value at K=5 is more optimal than the SI value at any other cluster. To analyze the data, the most optimal cluster validity value of K=5 is used.
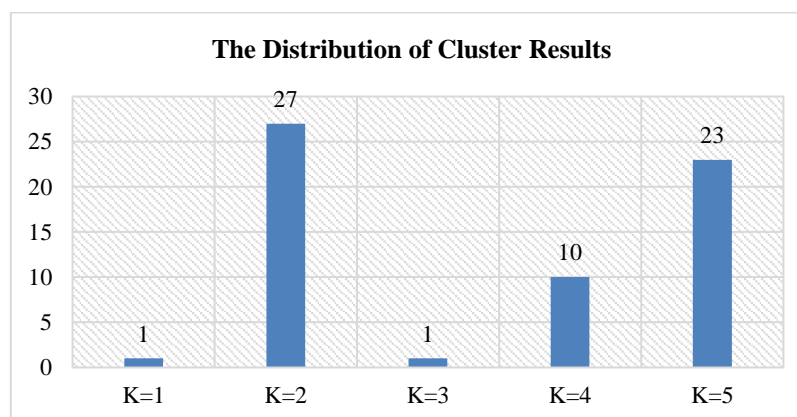

Figure 3. *Cluster Results Distribution*

Clustering at K=5 with a total of 62 students is presented in Figure 3. It was obtained that the distribution of student score data clusters with details K=2 is the cluster with the highest data of 27 students and K=1 and K=3 are clusters with the lowest data, which only contains 1 student in each cluster. From this study, it can be

seen that the Fuzzy C-Means algorithm can group student grade data into 5 clusters based on the best validity value even though the validity value is included in the weak structure category.

4. CONCLUSION

Based on the results and analysis carried out, it can be concluded that testing on mapping and grouping student value data by applying the Fuzzy C-Means algorithm obtained the best SI validity value from several cluster experiments, namely at K=5 of 0.4277. The distribution of K=5 in student score data with details is cluster 1 with a capacity of 1 student, cluster 2 with a capacity of 27 students, cluster 3 with a capacity of 1 student, cluster 4 with a capacity of 10 students, cluster 5 with a capacity of 23 students. The results were obtained by 2 groups based on the value of student learning outcomes on the data where students who had learning outcomes scores of less than 80 were in clusters 1 and 2. Students who have learning outcomes above 80 are in clusters 3, 4, and 5. The results of this study contribute to providing a better picture of the details of student learning outcomes and provide a basis for schools to conduct more effective coaching to improve overall student learning success. In addition, it is also necessary to conduct a more in-depth analysis of the factors that affect student learning outcomes at SMKN 5 Pekanbaru to provide more comprehensive insights for the school in designing a more effective coaching program. As recommendations for future research, it can be considered to apply other methods or algorithms that can improve the validity of results, as well as more deeply understand the factors and features that affect student learning outcomes in the school scope.

REFERENCES

[1] R. Maulana, N. Dwiyani, "Web Usage Mining Menggunakan K-Means untuk Mengetahui Kecenderungan Akses Pengguna (Studi Kasus: Ganto. Co)", *Jurnal Vocational Teknik Elektronika dan Informatika*, vol. 6, no. 2, 2018.

[2] J. Triwidianti, S. Yanto, and M. Prasojo, "EasyChair Preprint Comparison of Data Mining Methods for Predictive Achievement of Vocational Middle Education Students at State Vocational School (SMKN 1) Gadingrejo Pringsewu Lampung," 2021.

[3] W. Masrul, H. Silva, and T. Sundari, "Pelatihan Sketch Up Untuk Guru SMKN 5 Pekanbaru Pada Jurusan Desain Permodelan Dan Informasi Bangunan," *COMSEP: Jurnal Pengabdian Kepada Masyarakat*, vol. 3, no. 2, 2022.

[4] T. Nabillah and A. P. Abadi, "Prosiding Seminar Nasional Matematika dan Pendidikan Matematika Sesiomadika," 2019.

[5] D. Transaksi Bongkar Muat di Provinsi Riau, I. Kamila, U. Khairunnisa, P. Studi Sistem Informasi, and F. Sains dan Teknologi UIN Sultan Syarif Kasim Riau, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, pp. 119–125, 2019.

[6] H. J. Kim, J. W. Baek, and K. Chung, "Associative Knowledge Graph Using Fuzzy Clustering and Min-Max Normalization in Video Contents," *IEEE Access*, vol. 9, pp. 74802–74816, 2021, doi: 10.1109/ACCESS.2021.3080180.

[7] R. Sanjaya and Y. Nataliani, "Perbandingan Pembobotan Kriteria dan Seleksi Kriteria pada Pengelompokan Kinerja Karyawan dengan Fuzzy C-Means 1."

[8] A. Jaini, A. Weni Syaputri, T. Qurahman, S. Thaufik Rizaldi, P. H. Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau Jl Soebrantas Km, and P. Pekanbaru Riau, "Perbandingan Algoritma Fuzzy C-Means dan K-Medoids untuk Pengelompokan Data Penjualan pada 212 Mart," 2020.

[9] I. P. Sari, A.-K. Al-Khowarizmi, and I. H. Batubara, "Cluster Analysis Using K-Means Algorithm and Fuzzy C-Means Clustering For Grouping Students' Abilities In Online Learning Process," *Journal of Computer Science, Information Technology and Telecommunication Engineering*, vol. 2, no. 1, pp. 139–144, 2021.

[10] J. Nurkholis, K. Oktafianto, H. Thohari, P. Studi Matematika, F. Universitas PGRI Ronggolawe, and S. Negeri, "CLUSTERING DATA KENAIKAN KELAS SISWA MADRASAH TSANAWIYAH (MTs) MENGGUNAKAN METODE FUZZY C-MEANS (STUDI KASUS MTs PLUS AL AMIN BANJAREJO)," vol. 04, no. 01, pp. 11–18, 2022.

[11] I. Singh, S. Sabitha, T. Choudhury, A. Aggarwal, and B. K. Dewangan, "Mapping Student Performance With Employment Using Fuzzy C-Means," *International Journal of Information System Modeling and Design (IJISMD)*, vol. 11, no. 4, pp. 36–52, 2020.

[12] S. Sarah, "Analisis Penerimaan Vaksin Covid-19 Berbasis Fuzzy Clustering Machine Learning di Provinsi Riau," *Jurnal Riset Komputer)*, vol. 8, no. 6, pp. 2407–389, 2021, doi: 10.30865/jurikom.v8i6.3636.

[13]   H. S. Firdaus, A. L. Nugraha, B. Sasmito, M. Awaluddin, and C. A. Nanda, "PERBANDINGAN METODE FUZZY C-MEANS DAN K-MEANS UNTUK PEMETAAN DAERAH RAWAN KRIMINALITAS DI KOTA SEMARANG," 2021.

[14]   N. Ulinnuha, "Provincial Clustering in Indonesia Based on Plantation Production Using Fuzzy C-Means," *ITSMART: Jurnal Teknologi dan Informasi*, vol. 9, no. 1, pp. 8–12, 2020.

[15]   S. Kurniawan, A. Siregar, and H. Novita, "Penerapan Algoritma K-Means dan Fuzzy C-Means Dalam Mengelompokan Prestasi Siswa Berdasarkan Nilai Akademik," *Scientific Student Journal for Information, Technology and Science*, vol. 4, no. 1, pp. 73–81, 2023.

[16]   D. P. Sari, F. Sembiring, D. P. Sulisdianto, Y. Jentiner, and E. Suratno, "Implementasi Algoritma Fuzzy C-Means Untuk Memprediksi Faktor Siswa Membolos (Studi Kasus: Faktor Siswa Membolos Di Smpn 1 Parakansalak)," *Jurnal RESTIKOM: Riset Teknik Informatika dan Komputer*, vol. 2, no. 1, pp. 1–7, 2020.

[17]   K. Sinaga and K. Handoko, "IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI KELULUSAN SISWA DENGAN METODE NAÏVE BAYES," 2021.

[18]   F. A. Khan, K. Zeb, M. Al-Rakhami, A. Derhab, and S. A. C. Bukhari, "Detection and Prediction of Diabetes Using Data Mining: A Comprehensive Review," *IEEE Access*, vol. 9. Institute of Electrical and Electronics Engineers Inc., pp. 43711–43735, 2021. doi: 10.1109/ACCESS.2021.3059343.

[19]   Z. Li, C. Tang, X. Liu, X. Zheng, W. Zhang, and E. Zhu, "Consensus Graph Learning for Multi-View Clustering," *IEEE Trans Multimedia*, vol. 24, pp. 2461–2472, 2022, doi: 10.1109/TMM.2021.3081930.

[20]   S. Fitri Octavia, "Penerapan K-Means dan Fuzzy C-Means untuk Pengelompokan Data Kasus Covid-19 di Kabupaten Indragiri Hilir," *Technology and Science (BITS)*, vol. 3, no. 2, 2021, doi: 10.47065/bits.v3i2.1005.

[21]   M. H. Qasem, N. Obeid, A. Hudaib, M. A. Almaiah, A. Al-Zahrani, and A. Al-Khasawneh, "Multi-Agent System Combined with Distributed Data Mining for Mutual Collaboration Classification," *IEEE Access*, vol. 9, pp. 70531–70547, 2021, doi: 10.1109/ACCESS.2021.3074125.

[22]   I. Irwan, S. Sidjara, and A. P. Aryati, "Pengelompokan Jenis Penerimaan Pajak di Kota Makassar Menggunakan Fuzzy Clustering," *Euler : Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 10, no. 1, pp. 98–102, May 2022, doi: 10.34312/euler.v10i1.14225.

[23]   D. S. Rohmah, D. Retno, and S. Saputro, "Clustering Data dengan Algoritme Fuzzy c-Means Berbasis Indeks Validitas Partition Coefficient and Exponential Separation (PCAES). PRISMA, Prosiding Seminar Nasional Matematika," vol. 3, pp. 58–63, 2020, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/

[24]   D. Fitrianah, W. Gunawan, and R. Algian Kurniaputra, "Techno Xplore Jurnal Ilmu Komputer dan Teknologi Informasi Implementasi Algoritma DBScan dalam Pemngambilan Data Menggunakan Scatterplot."

[25]   Mustakim, M. Z. Fauzi, Mustafa, A. Abdullah, and Rohayati, "Clustering of Public Opinion on Natural Disasters in Indonesia Using DBSCAN and K-Medoids Algorithms," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1742-6596/1783/1/012016.

[26]   G. Ghale, "Peningkatan Hasil Belajar IPA Pada Siswa Kelas VI SDK Wologeru Melalui Model POE (Predict Observed Explain)."

[27]   P. Studi and T. Informatika, "DATA MINING UNTUK MEMPREDIKSI PRESTASI MAHASISWA INSTITUT BISNIS DAN INFORMATIKA KOSGORO 1957 BERDASARKAN SOSIAL EKONOMI ORANG TUA DAN PRESTASI MASA LALU," vol. 2, no. 2, pp. 60–71, 2021.

[28]   S. Anwar, T. Suprapti, G. Dwilestari, I. Ali, P. Studi Rekayasa Perangkat Lunak Jln Perjuangan No, and B. Kesambi Kota Cirebon, "PENGELOMPOKKAN HASIL BELAJAR SISWA DENGAN METODE CLUSTERING K-MEANS Program Studi Sistem Informasi Jln Perjuangan No 10B Kesambi Kota Cirebon 4)," *Jurnal Sistem Informasi dan Teknologi Informasi)*, vol. 4, no. 2, pp. 60–72, 2022.

[29]   B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "ANALISIS DATA MINING DATA NETFLIX MENGGUNAKAN APLIKASI RAPID MINER," *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 1, Apr. 2021, doi: 10.30813/jbase.v4i1.2729.

[30]   G. Seroklinov and A. Goonko, "Comparative analysis of experimental data clustering in MATLAB and Python environment," in *E3S Web of Conferences*, EDP Sciences, 2023, p. 02025.