Personalized Student Learning Mechanisms using K-Means and K-Medoids Clustering Algorithms based on Individual Preferences

Rendy Wenda Dwi Kurniawan
University of Technology Yogyakarta
North Ring Road, Jombor, Sleman, DIY, Indonesia

Arief Hermawan
University of Technology Yogyakarta
North Ring Road, Jombor, Sleman, DIY, Indonesia

ABSTRACT
This research aims to minimize the mismatch between competencies and career interests, a common issue faced by students who often find themselves in the wrong field of study or working in jobs unrelated to their educational background. Determining an effective and efficient learning mechanism through precise student clustering to form optimal learning groups is an effort that can be made to better align with the individual preferences of each student. However, the process of clustering high school students encounters challenges such as resource constraints, time limitations, and the need for effective outcomes. Therefore, this research explores a solution by implementing a clustering process for high school students based on individual preferences, including subject interests, career aspirations, and preferred learning methods, using the K-Means and K-Medoids algorithms. Performance analysis of both algorithms reveals that K-Means outperforms in handling student preference data, resulting in an optimal number of clusters of 6. The model evaluation results indicate a Silhouette Coefficient of 0.786 for K-Means and a Davies-Bouldin Index of 0.334.

Keywords
Clustering, K-Means, K-Medoids, Learning Mechanisms, Individual Preferences.

1. INTRODUCTION
Students are integral components of the education process, serving as the focal point of learning with the goal of continuous development towards becoming high-quality individuals. In pursuit of educational goals, achieving the effectiveness and efficiency of the learning process is highly essential. Therefore, optimization of the learning mechanisms being implemented needs to be continuously improved and tailored, especially for high school students who will either proceed to higher education or directly enter the professional world. According to the results of a research survey conducted by the Indonesia Career Center Network (ICCN), 87% of students acknowledge that the majors they have chosen do not align with their interests, and 71% of professionals work in fields unrelated to their educational background[1]. According to statements made by the Institute for Development of Economics and Finance (INDEF), 60.62% indicate a mismatch between their job field and their educational background and skills[2]. The Minister of Education, Culture, Research, and Technology, Nadiem Makarim, has expressed a similar concern, stating that 80% of students do not work in fields related to their study programs or majors[3]. Referring to these statements, to minimize the inadequacy or mismatch of competencies in relation to career interests, it can be achieved through the determination of effective and efficient learning mechanisms, especially at the high school level. The process of determining effective and efficient learning mechanisms can begin by appropriately clustering students to form optimal study groups or sets.

In the process of clustering high school students to form optimal study groups, various challenges are often encountered. These optimal study groups can be defined as a group of students who share common interests and goals. However, when the clustering process is solely based on academic performance, it may not yield optimal results, especially in the context of high school students who are transitioning to higher education or entering the professional world. Considering the prevalence of academic misconduct, such as cheating and the subjectivity in grading by educators, relying solely on academic performance may not be the most effective approach. Therefore, to create optimal study groups, it is necessary to perform clustering of high school students based on their individual preferences, such as preferred subjects, career aspirations, and favored learning methods. This approach serves as a solution to address the mismatch or lack of alignment between students' interests and skills with their future career paths. Because people have different learning styles, it is crucial to create and adapt the learning mechanisms to student preferences to maximize and accelerate the educational process[4]. By clustering students based on their individual preferences, it is expected that the learning mechanisms can be better tailored to each student's career interests. This approach aims to enhance the alignment between students' academic experiences and their future career aspirations.

Referring to [5], this research discusses the grouping of student data in advanced classes, facilitating tailored education within the school environment based on individual abilities. Utilizing the K-Means algorithm, data clustering is performed based on variables such as competence, skill grades, and software proficiency. In this study, the K-Means algorithm proves to be effective in addressing student data grouping challenges. Furthermore, as referenced in[6], a study investigates student clustering for non-formal education selection using variables such as practical exam scores, final exam grades, and interview scores. In this research, the K-Medoids algorithm is employed and found to be suitable for student clustering. Building upon the aforementioned references, the clustering process for high school students based on individual preferences will utilize both the K-Means and K-Medoids algorithms. Consequently, an analysis will be conducted to assess the outcomes of applying these algorithms to cluster high school students based on their individual preferences. K-Means is an unsupervised learning algorithm utilized for data partitioning into multiple clusters represented by centroids as the centers of each cluster.

By employing this algorithm, data with high similarity are grouped into the same cluster with the aim of minimizing variation within clusters while maximizing it between clusters. On the other hand, K-Medoids is a technique in cluster analysis used to group data by identifying a set of medoids, which are data points representing the center of each cluster. Medoids are data points within a cluster that have the minimum average distance to all other data points in the same cluster. This algorithm differs from K-Means, which employs centroids as cluster representations. From the use of both the K-Means and K-Medoids algorithms, an analysis will be conducted to assess the outcomes obtained. This is done with the aim of determining which algorithm is better suited to the data and the issues at hand, which involve the clustering of high school students as an effort to align the learning mechanisms with career interests based on individual preferences.

2. PROPOSED APPROACH

In the implementation of research related to the clustering of high school students as an effort to align learning mechanisms with career interests based on individual preferences, several stages are undertaken. These research stages have been depicted as a conceptual design that outlines the fundamental workflow through a flowchart as follows.

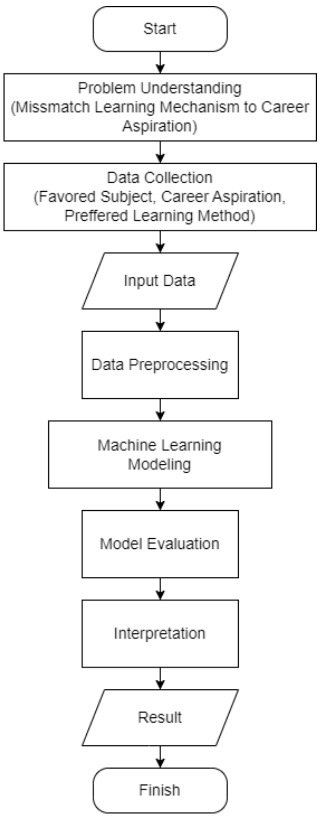

Fig 1: Flowchart System

1. Problem Understanding

The lack of adequate preparation and fundamental competence can be a contributing factor to the mismatch between the chosen career path and the educational background. Therefore, in the clustering process of high school students who are transitioning to college or entering the workforce directly, it is crucial to be carried out accurately. This is done to establish optimal learning groups as an effort to adjust the learning mechanisms to career interests, based on individual preferences.

2. Data Collection

The data for this research was obtained through a survey approach, involving the collection of information via questionnaires regarding the interests and preferences of high school students, who are the focus of individual preferences, at one of the high schools in East Java, specifically SMAN 1 Karangjati. In this activity, questionnaires were distributed to students with inquiries related to their favorite subjects, desired professions, and preferred learning methods. The questionnaires were completed using one of the services provided by Google, Google Forms, and the results were stored in CSV format.

3. Input Data

The data collected through the survey, encompassing questions concerning the individual preferences of high school students, will be used in the research on clustering that utilizes machine learning models using the Python programming language.

4. Data Preprocessing

The data related to the individual preferences of high school students obtained from the survey is categorical in nature. Therefore, a preprocessing step is required to convert each value into numerical form to enable its use in subsequent processes, specifically, the clustering process using machine learning models.

5. Machine Learning Modeling

The creation of the machine learning model utilizes the K-Means and K-Medoids algorithms with the Python programming language. The results from the application of these two machine learning models will undergo an analytical process to determine which algorithm is more suitable and effective given the data and challenges at hand.

6. Model Evaluation

In the course of the analysis, to identify the algorithm that is better suited to the data and the challenges faced, a model evaluation step is crucial to assess the performance of both machine learning models. This model evaluation step will employ the Silhouette coefficient and the Davies-Bouldin Index.

7. Interpretation

Following the clustering results obtained, understanding and interpreting the characteristics of each formed group becomes essential. This interpretation process is necessary to facilitate the practical utilization of the clustering results achieved through machine learning as a solution to the challenges at hand.

3. METHOD

3.1 K-Means

K-Means is a clustering algorithm used to partition data into multiple clusters. This algorithm partitions data into a predefined number of clusters (K) with the aim of achieving high similarity within each cluster while maintaining low similarity between clusters[7]. K-Means is one of the most commonly used algorithms in the data clustering process. This is due to its ease of implementation, relatively short processing time for clustering, strong adaptability, and suitability for handling a large number of data points and clusters[8]. Referring to the definitions previously presented, K-Means is an unsupervised learning algorithm that performs well in partitioning data into distinct clusters. The goal is to group similar data points into the same cluster to minimize variation within a cluster and maximize it between clusters.

The steps in using the K-Means clustering algorithm can be explained as follows[9]:

1. Input the data.
2. Determine the total number of clusters.
3. Select data randomly to serve as initial centroids.
4. Calculate the nearest distance between each data point and the centroids using the following formula

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (3.1)$$

Where:
d = Distance
x1 = Latitude coordinate of the data
x2 = Latitude coordinate of the centroid
y1 = Longitude coordinate of the data
y2 = Longitude coordinate of the centroid

5. Calculate the center of each cluster with the new cluster members.
6. The clustering process is considered complete if the cluster centers do not change. However, if the cluster centers continue to change, repeat the distance calculation steps until the cluster centers no longer change.

3.2 K-Medoids

The K-Medoids algorithm, also known as Partitioning Around Medoids (PAM), was developed by Kaufman and Rousseuw in 1987. K-Medoids is a non-hierarchical partition-based clustering method that employs the medoid as the center of each cluster. The algorithm commonly used in K-Medoids is Partitioning Around Medoids (PAM). This algorithm aims to minimize the distance between objects and the medoid as the center of the cluster[10]. The initial step of this algorithm involves selecting a k-medoid, which is then exchanged with a non-medoid object. This process aims to enhance the quality of the clustering. The K-Medoids algorithm exhibits greater robustness compared to K-Means, especially when dealing with datasets containing extreme values or outliers. In this algorithm, the object designated as the representative of the cluster is the medoid, as opposed to the mean value of objects within a cluster based on a reference point[11]. Referring to the various literature sources as previously discussed, K-Medoids algorithm is a data clustering method aimed at partitioning data into multiple clusters, each with a central point known as a medoid, representing the most typical object in that cluster. The objective is to minimize the distance between data objects and their respective medoids within the corresponding cluster, thereby achieving homogeneity within the clusters.

Here are the stages in the K-Medoids algorithm, as referenced in[12].

1. Initialization: Randomly select K data objects as initial medoids. K represents the pre-defined number of clusters.
2. Assignment: Each data object is assigned to the nearest medoid based on distance or similarity.
3. Evaluation: Calculate the total distance or similarity between each object and its medoid within the cluster.
4. Swap: Attempt to swap a medoid with a non-medoid object in the same cluster if it improves cluster quality.
5. Iteration: Repeat steps 2-4 until no further changes enhance cluster quality.
6. The final result is a set of K-medoids that produce optimal clusters based on the relevant distance or similarity measure.

3.3 Elbow Method

The Elbow method is a technique used to obtain information for determining the optimal number of clusters. This is achieved by assessing the percentage of variance explained as the number of clusters increases and identifying a point where an 'elbow' becomes apparent[13]. The elbow method calculates the squared differences for various K values. As K increases, the average distortion level decreases. The number of samples within each category decreases, and the samples get closer to the center of gravity. As K increases, the point at which the improvement in the degree of distortion decreases most significantly is the K value corresponding to the elbow[14]. The Elbow Method aids in determining the optimal number of clusters by identifying a point where the rate of variance or distortion reduction sharply changes, indicating the most suitable cluster count. Based on the provided definitions, The Elbow Method aids in determining the optimal number of clusters by identifying a point where the rate of variance or distortion reduction sharply changes, indicating the most suitable cluster count.

3.4 Silhouette Coefficient

The Silhouette Index (SI), or silhouette coefficient, is commonly utilized to assess the quality and strength of clusters, particularly in terms of how well an object is positioned within a cluster. This method is employed for the validation of individual data points, single clusters, or even the entirety of clusters. It is frequently employed in the validation of clusters by combining cohesion and separation values[15]. The advantage of SC depends solely on the partition of the dataset, rather than on the clustering algorithm itself[16]. Therefore, it can be concluded that, the Silhouette Coefficient assesses the quality of clusters by measuring how well each data point aligns with its assigned cluster, with values closer to 1 signifying better cluster cohesion.

3.5 Davies-Bouldin Index

The Davies-Bouldin Index (DBI) is one of the methods used to measure cluster validity in a clustering method. The measurement with the Davies-Bouldin Index aims to maximize the distance between clusters while simultaneously minimizing the distance between points within a cluster[17]. The smaller the Davies-Bouldin Index (DBI) obtained (non-negative >= 0), the better the quality of the produced clusters[18]. From the several definitions that have been presented, The Davies-Bouldin Index gauges cluster quality by considering both inter-cluster separation and intra-cluster cohesion, with lower values indicating more distinct and compact clusters.

4. RESULTS AND DISCUSSION

4.1 Dataset

The research involving the analysis of the application of the K-Means and K-Medoids algorithms in clustering students based on individual preferences was conducted using data obtained through a questionnaire survey involving 161 student respondents. The questionnaire encompassed variables related to preferred subjects, desired occupations, and favored learning methods. The initial data collected were in categorical form, requiring the conversion of values into numerical representations. This transformation process was carried out manually by establishing specific indicators for each category within each question. The stages of data collection and indicator determination were informed by the insights and analysis of local school teachers, as well as the author's expertise, which were adapted from a variety of sources covering a wide range of occupational fields and the corresponding subject matter. The statistical properties of each variable within the dataset can be observed through the following histograms.

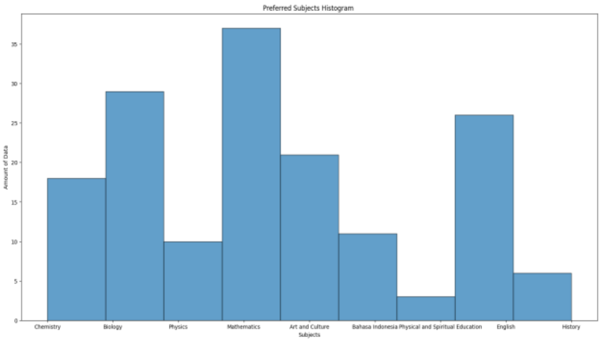

Fig 2: Preferred Subjects Histogram

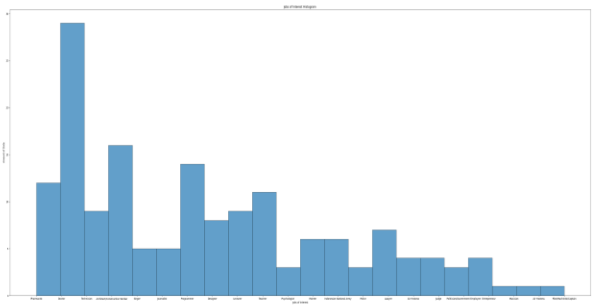

Fig 3: Jobs of Interest Histogram

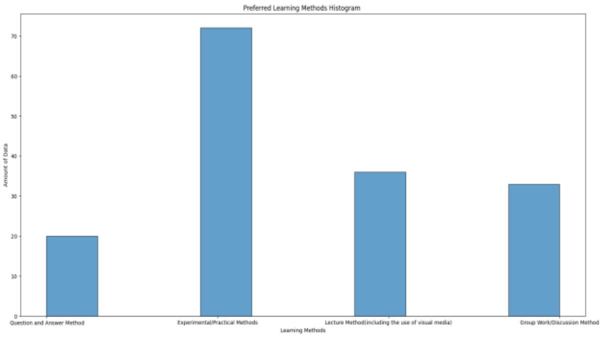

Fig 4: Preferred Learning Methods Histogram

From the visualizations of each variable above, it is evident that the data obtained from the questionnaire survey show an uneven distribution regarding student preferences. Among the variables, Mathematics is the most favored subject. The occupation with the highest level of interest is that of a doctor. Meanwhile, the most favored learning method is experimentation or practical application.

Based on the data presented in the explanations and the visualizations above, the distribution of each variable can be observed in the form of a 3D scatter plot. In a 3D scatter plot, data is represented as points in three-dimensional space, with each axis representing one variable. This representation is essential for data analysis, as it enables the direct identification of structures, correlations, or patterns among the three variables under investigation, namely, preferred subjects, desired occupations, and favored learning methods. The outcomes of the 3D scatter plot visualization of student preferences are presented as follows.

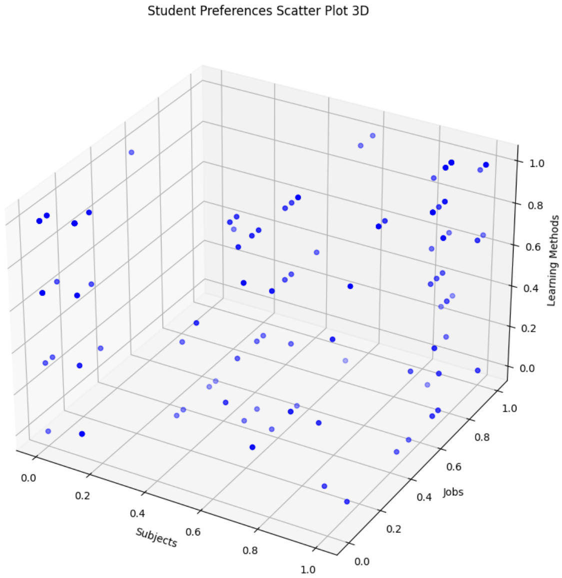

Fig 5: Student Preferences Scatter Plot 3D

4.2 Implementation of K-Means

In K-means clustering modeling, the first step involves utilizing the elbow method to assist in determining the optimal number of clusters. When combining the elbow method with K-means, the optimal number of clusters is indicated by the presence of an elbow point formed due to a significant change in inertia. The optimal number of clusters resulting from the combination of the elbow method with K-means is 2, 3, 5, and 6. To determine the best cluster count, further model evaluation is required, which will be discussed in the following section. The visualization results of the elbow method combined with K-means can be observed as follows.

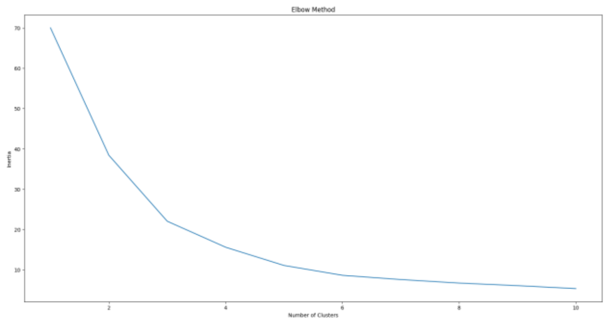

Fig 6: Elbow Method for K-Means

4.3 Implementation of K-Medoids

The clustering experiment was conducted once more, but with a different algorithm, namely K-medoids. As an initial step, the determination of the optimal number of clusters was performed by combining the elbow method and the K-medoids algorithm. Just as explained above, the optimal number of clusters is marked by the presence of an elbow point resulting from a significant change in inertia. The application of the elbow method combined with K-medoids yielded quite different results. The combination of the elbow method and K-medoids indicated that the optimal number of clusters is 3, 4, and 6. From these two sets of results, it can be concluded that the application of the elbow method, when combined with both K-means and K-medoids, yields similar results for the determination of the optimal number of clusters, which are 3 and 6. The visualization results of the elbow method and K-medoids can be observed as follows.

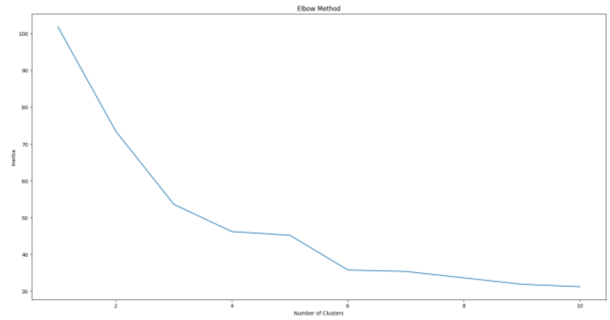

Fig 7: Elbow Method for K-Medoids

4.4 Model Evaluation

The modeling process using the K-means and K-medoids algorithms has been completed with the determination of the optimal number of clusters through the elbow method. From the modeling results, it is necessary to perform testing to evaluate the obtained clustering outcomes. In this study, the model evaluation process is conducted using the Silhouette Coefficient and Davies-Bouldin Index parameters. In the assessment, if the Silhouette Coefficient approaches 1, the clustering results with that number of clusters are considered better. Meanwhile, for the Davies-Bouldin Index, clustering outcomes are deemed superior if the value approaches 0. The visualization results of the application of the Silhouette Coefficient and Davies-Bouldin Index for each model can be seen below.

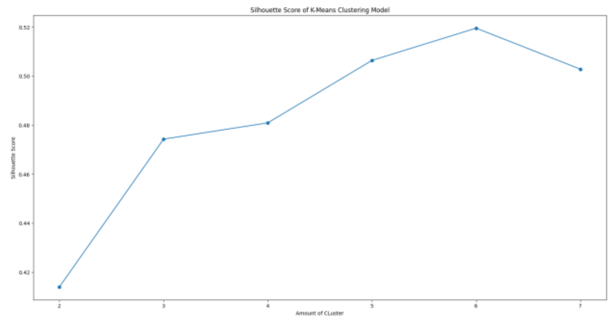

Fig 8: Silhouette Score of K-Means Model

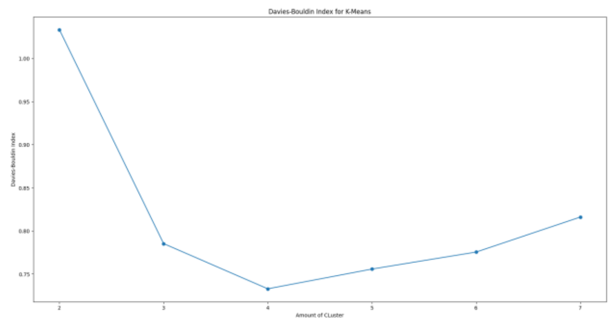

Fig 9: Davies-Bouldin Index of K-Means Model

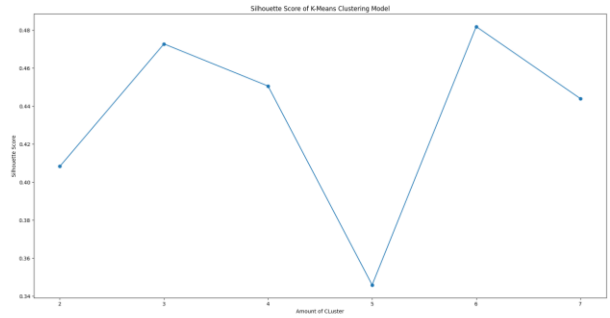

Fig 10: Silhouette Score of K-Medoids Model

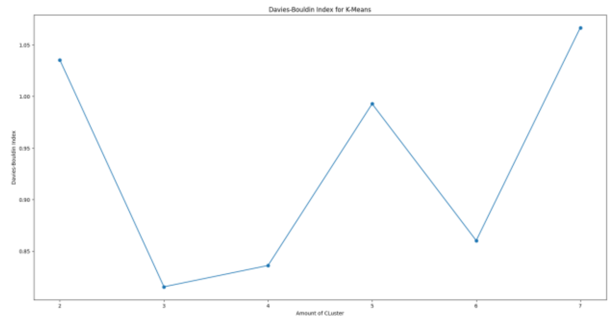

Fig 11: Davies-Bouldin Index of K-Medoids Model

Based on the results of the visualizations using the Silhouette Coefficient Score and the Davies-Bouldin Index, several conclusions can be drawn. In the application of the K-Means and K-Medoids methods, the evaluation of models using the Silhouette Coefficient reveals that the optimal number of clusters for data clustering is the same for both methods, which is 6. Similarly, when evaluating using the Davies-Bouldin Index, the best number of clusters for data clustering using both the K-Means and K-Medoids algorithms is also 6. This is because in determining the optimal number of clusters using the Davies-Bouldin Index, it is not only based on the lowest point but also takes into consideration the elbow produced in the graph. Therefore, the number of clusters that yield the best data clustering, based on the determination using the elbow method and the model evaluation conducted, is 6. Considering the comparison of the evaluation results, the algorithm that performs better and is suitable for the data and the problem related to adjusting the learning mechanism based on individual preferences is K-Means. For a more detailed comparison of the model evaluations for the K-Means and K-Medoids algorithms, please refer to the table below:

Table 1. Model Evaluation Results Comparison

Algorithm	Cluster	SI	DBI
K-Means	3	0.679	0.510
K-Means	4	0.732	0.396
K-Means	**6**	**0.786**	**0.334**
K-Medoids	3	0.668	0.541
K-Medoids	4	0.701	0.450
K-Medoids	**6**	**0.770**	**0.340**

4.5 Interpretation

Based on the various stages conducted previously, from problem understanding to model evaluation, the clustering results using K-Means as the algorithm demonstrate strong performance and alignment with the data and the problem at hand. Out of the 161 data points, they have been effectively grouped into six clusters, which is deemed to be the most optimal number. The outcomes obtained from implementing K-Means with six clusters as the superior algorithm can be observed in both visual and tabular formats, as presented below

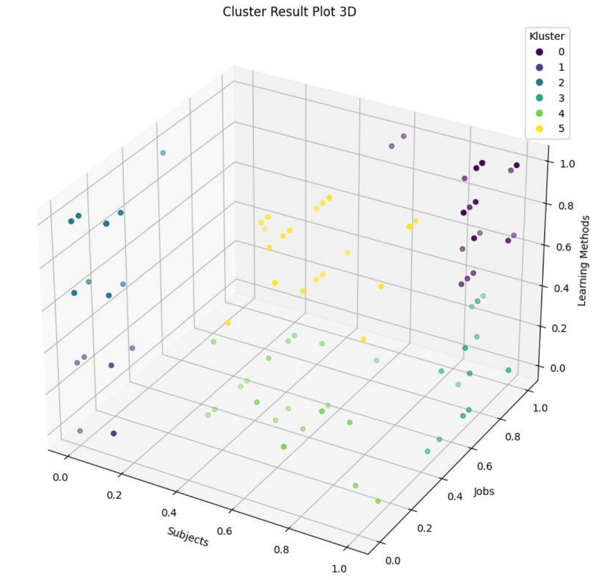

Fig 12: Cluster Result Scatter Plot 3D

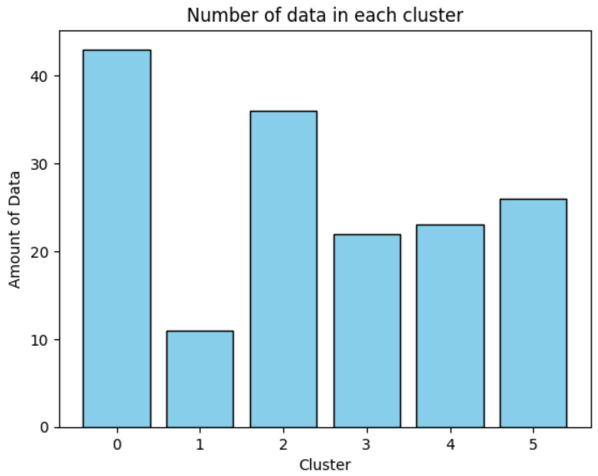

Fig 13: Number of Data in each Cluster

Based on the clustering results up to the interpretation stage, it is anticipated that these findings can be applied in real-life situations. Adjusting the learning mechanisms and determining appropriate steps based on students' individual preferences regarding career interests can provide a solution to the underlying issues addressed in this research. Consequently, high school-level students can better prepare themselves for further education at the tertiary level or for immediate entry into the workforce.

Table 2. Cluster Labels

Cluster	Number of Data	Cluster Labels
0	43	Innovative Engineering Explorers
1	11	Medical Sciences Enthusiasts
2	36	Health Innovators and Practitioners
3	22	Economic and Business Scholars
4	23	Legal Scholars and Governance Analysts
5	26	Educational Language Discourse

In the cluster analysis results, the cluster "Innovative Engineering Explorers" demonstrates a strong interest in engineering and technology. Students in this cluster tend to explore engineering concepts through practical and experimental approaches. In addressing these findings, it is recommended to provide facilities and equipment for practical engineering experiments and projects. These steps involve encouraging students to develop innovative projects, organizing exhibitions, or presenting their work. Additionally, providing additional resources to support the exploration of engineering concepts beyond the classroom environment is crucial. Meanwhile, in the "Medical Sciences Enthusiasts" cluster, students show a specific interest in health sciences and medicine. They are inclined towards practical application and questions/responses related to health content. Managing this cluster involves recommending opportunities for field visits or practical experiences in local healthcare facilities. Encouraging students to ask questions, participate in group discussions, present their medical findings, and facilitating question-and-answer sessions with medical experts or healthcare practitioners is essential.

Moving on to the "Health Innovators and Practitioners" cluster, which encompasses an interest in Health and Medical Sciences along with experiments/practice. The term "Innovators and Practitioners" reflects a combination of health expertise and practical application in the field. To address these results, it is suggested to form collaborative projects that merge health concepts with practical applications. Encouraging students to develop innovations in healthcare through research and experiments, as well as providing opportunities for internships or practical experiences in local healthcare environments, is recommended. Next, the "Economic and Business Scholars" cluster demonstrates a specific interest in economics and business, with an emphasis on lectures (including visual media) as the primary learning method. To address these findings, it is advisable to use teaching methods that utilize visual materials and business case studies to enhance understanding. Organizing seminars or inviting guest lectures from economics and business professionals, as well as assigning projects or tasks that connect economic theories with real-world business applications, is recommended.

In the "Legal Scholars and Governance Analysts" cluster focuses on law and governance, primarily utilizing lectures (including visual media) as the main learning approach. Managing this group involves organizing discussion sessions and analyzing legal case studies to foster in-depth understanding. Inviting legal experts or governance practitioners as guest speakers and developing legal projects or simulations that actively involve students are also recommended. Lastly, the "Educational Language Discourse" cluster students with a deep interest in education and language. This cluster actively engages in discussions and group work to deepen their understanding of educational and language concepts. Addressing these results involves advocating for more discussion and group work sessions as an integral part of the learning process. These steps include facilitating group projects, using educational materials that support discussions, and implementing teaching methods that encourage active student participation.

5. CONCLUSION

Referring to each stage of discussion related to the analysis of the application of the K-Means and K-Medoids algorithms in the effort to tailor the learning mechanism based on individual preferences, several conclusions can be drawn. Both the K-Means and K-Medoids algorithms exhibit strong performance in clustering student preference data. However, the evaluation results indicate that K-Means outperforms K-Medoids. This can be observed from the model evaluation using the silhouette coefficient, which shows the best number of clusters for each algorithm, both being 6, with K-Means achieving a score of 0.786 and K-Medoids scoring 0.770. Furthermore, based on the model evaluation using the Davies-Bouldin index, K-Means obtained a score of 0.334, while K-Medoids scored 0.340.From the results of this model evaluation, it can be discerned that the K-Means algorithm is superior in handling student preference data, and the optimal number of clusters is 6. With the clustering process using the superior algorithm and the optimal number of clusters, an interpretation of the characteristics of each cluster has been obtained. This interpretation is expected to facilitate the utilization of this information as decision support in determining policies related to learning mechanisms based on student preferences.

6. REFERENCES

[1] Gumanti Awaliyah and Dwi Murdaningsih, "87 Persen Mahasiswa Mengaku Salah Pilih Jurusan," Republika. [Online]. Available: https://republika.co.id/berita/pmjuhw368/87-persen-mahasiswa-mengaku-salah-pilih-jurusan

[2] Lalu Rahadian, "Skill Tak Sesuai Suplai Tenaga Kerja Tak Terserap," Bisnis.com. [Online]. Available: https://ekonomi.bisnis.com/read/20190316/12/900380/skill-tak-sesuai-suplai-tenaga-kerja-tak-terserap

[3] Novia Aisyah, "Nadiem Ungkap 80% Lulusan Tak Bekerja Sesuai Prodi, Bagaimana Sisanya?," detikEdu. [Online]. Available: https://www.detik.com/edu/perguruan-tinggi/d-5793585/nadiem-ungkap-80-lulusan-tak-bekerja-sesuai-prodi-bagaimana-sisanya

[4] A. Moubayed, M. Injadat, A. Shami, and H. Lutfiyya, "Student Engagement Level in an e-Learning Environment: Clustering Using K-means," American Journal of Distance Education, vol. 34, no. 2, pp. 137–156, Apr. 2020, doi: 10.1080/08923647.2020.1696140.

[5] J. Hutagalung, Y. Hendro Syahputra, Z. Pertiwi Tanjung, S. Triguna Dharma, and J. I. Pintu Air, "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering," Hal AH Nasution, vol. 9, no. 1, 2022, [Online]. Available: http://jurnal.mdp.ac.id

[6] N. Saurina, L. Retnawati, F. H. Sukma Pratama, and U. Pudjianto, "Pengelompokan Seleksi Siswa Baru di Lembaga Pendidikan Non Formal Kabupaten Gresik Menggunakan Clustering K-Medoids," Jurnal Sains dan Informatika, pp. 36–45, Jun. 2023, doi: 10.34128/jsi.v9i1.527.

[7] V. R. Reddy, D. Yakobu, S. Shiva Prasad, and P. A. Harsha Vardhini, "Clustering Student Learners based on performance using K-Means Algorithm," in 2022 International Mobile and Embedded Technology Conference, MECON 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 302–306. doi: 10.1109/MECON53876.2022.9752165.

[8] N. Hidayat, R. Wardoyo, U. Gadjah Mada, I. S. Azhari, and H. Dwi Surjono, "Enhanced Performance of the Automatic Learning Style Detection Model using a Combination of Modified K-Means Algorithm and Naive Bayesian," 2020. [Online]. Available: www.ijacsa.thesai.org

[9] Z. M. Ali, N. H. Hassoon, W. S. Ahmed, and H. N. Abed, "The Application of Data Mining for Predicting Academic Performance Using K-means Clustering and Naïve Bayes Classification," International Journal of Psychosocial Rehabilitation, vol. 24, no. 03, pp. 2143–2151, Feb. 2020, doi: 10.37200/ijpr/v24i3/pr200962.

[10] S. Saket and S. Pandya, "Implementation of Extended K-Medoids Algorithm to Increase Efficiency and Scalability using Large Datasets," 2016.

[11] C. Oktarina, K. Anwar Notodiputro, and ‡ Indahwati, "COMPARISON OF K-MEANS CLUSTERING METHOD AND K-MEDOIDS ON TWITTER DATA *," 2020.

[12] N. Qona'ah, A. Rachma Devi, I. Made, and G. M. Dana, "Laboratory Clustering using K-Means, K-Medoids, and Model-Based Clustering," 2020.

[13] R. Nainggolan, R. Perangin-Angin, E. Simarmata, and A. F. Tarigan, "Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method," in Journal of Physics: Conference Series, Institute of Physics Publishing, Dec. 2019. doi: 10.1088/1742-6596/1361/1/012015.

[14] M. Cui, "Introduction to the K-Means Clustering Algorithm Based on the Elbow Method", doi: 10.23977/accaf.2020.010102.

[15] D. T. Dinh, T. Fujinami, and V. N. Huynh, "Estimating the Optimal Number of Clusters in Categorical Data Clustering by Silhouette Coefficient," in Communications in Computer and Information Science, Springer Science and Business Media Deutschland GmbH, 2019, pp. 1–17. doi: 10.1007/978-981-15-1209-4_1.

[16] H. B. Tambunan, D. H. Barus, J. Hartono, A. S. Alam, D. A. Nugraha, and H. H. H. Usman, "Electrical peak load clustering analysis using K-means algorithm and silhouette coefficient," in Proceeding - 2nd International Conference on Technology and Policy in Electric Power and Energy, ICT-PEP 2020, Institute of Electrical and Electronics Engineers Inc., Sep. 2020, pp. 258–262. doi: 10.1109/ICT-PEP50916.2020.9249773.

[17] B. Jumadi Dehotman Sitompul, O. Salim Sitompul, and P. Sihombing, "Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm," in Journal of Physics: Conference Series, Institute of Physics Publishing, Jul. 2019. doi: 10.1088/1742-6596/1235/1/012015.

[18] H. Santoso and H. Magdalena, "Improved K-Means Algorithm on Home Industry Data Clustering in the Province of Bangka Belitung," in Proceeding - ICoSTA 2020: 2020 International Conference on Smart Technology and Applications: Empowering Industrial IoT by Implementing Green Technology for Sustainable Development, Institute of Electrical and Electronics Engineers Inc., Feb. 2020. doi: 10.1109/ICoSTA48221.2020.1570598913.