ISSN 2085-4552

Identifying Academic Performance Patterns Among PTIK Students Using K-Means Clustering

Rizak Al Hasbi Anwar[1], Febri Liantoni[2], Dwi Maryono[3]
[1,2,3] Dept. of Informatics Education, Sebelas Maret University, Indonesia
[1]rizakalhasbi_22@student.uns.ac.id, [2]Febri.liantoni@gmail.com, [3]dwimaryono@staff.uns.ac.id

Accepted 23 June 2025
Approved 06 January 2026

***Abstract*— This study explores the identification of academic performance patterns among students in the Informatics and Computer Engineering Education Study Program (PTIK) at Sebelas Maret University, focusing on the 2022 cohort. Using the K-Means clustering method within the scope of Data Mining, this research analyzes student performance data across multiple course categories from the first to fourth semesters. Through the Elbow method, four optimal clusters were established, each representing distinctive patterns of academic achievement. The analysis was conducted using RapidMiner software to reveal nuanced insights into student learning outcomes. Cluster 1 consists of students with moderate achievements in most categories, with a particular strength in Multimedia. Cluster 2 includes students with generally lower academic performance but shows a relative strength in General Courses. Cluster 3 is composed of high-achieving students who excel across categories, particularly in Software Engineering (RPL), Multimedia, and Educational subjects, indicating well-rounded academic proficiency. Cluster 4 comprises students with notable strengths in Software Engineering and Computer Networking, yet demonstrates lower performance in certain specialized subjects. These findings highlight the potential to tailor educational programs to address the specific learning needs and strengths of each student group, facilitating more personalized and effective academic support.**

Index Terms*— Clustering; K-Means; Data Mining; Academic Performance; RapidMiner.

I. Introduction

In the rapidly evolving field of Informatics and Computer Engineering, understanding student performance patterns has become increasingly important. Accurate identification of academic achievements and potential areas of improvement among students can significantly enhance educational strategies and contribute to more personalized learning experiences. Academic institutions are now leveraging data mining techniques to identify hidden trends in student performance data, which can help develop targeted interventions and improve overall learning outcomes. The increasing availability of student data has driven the adoption of data mining techniques to support academic decision-making and personalize learning [1]. The advancement of data mining technologies has influenced many researchers to investigate more profound insights into the knowledge dissemination process [2]. The application of data mining methods in the field of education has attracted great attention in recent years [3]. Data mining will certainly be very useful to analyze the activities of students who succeed and those who are at risk of failing, to develop improvement strategies based on students' academic performance, and therefore to assist educators in the development of pedagogical methods [4]. One such technique, K-Means clustering, has shown promise in grouping students based on similar characteristics, allowing for a detailed analysis of their academic tendencies and strengths. The K-Means Clustering algorithm is a widely used data analysis tool that divides data into many clusters according to shared attributes [5]. With the emergence of Big Data and the increased availability of educational datasets, clustering techniques can now be applied to identify distinct groups of students based on their academic performance across multiple subjects. These techniques are increasingly utilized in higher education to uncover hidden patterns in student achievement, enabling more effective academic planning and resource allocation [6].

However, in the context of higher education in Indonesia, particularly in programs such as Informatics and Computer Engineering Education (PTIK), there remains a lack of comprehensive, data-driven analysis regarding student academic achievement patterns. Most academic evaluations still rely on cumulative GPA or individual course performance without identifying group-based trends that can inform strategic interventions. Identifying these patterns is crucial because it allows institutions to detect underperforming groups early, tailor pedagogical approaches, and provide more equitable academic support [7]. Moreover, with the increasing complexity of interdisciplinary courses within PTIK, spanning software engineering, multimedia, and education science, students often exhibit diverse capabilities across subject categories. Without a structured analysis, such diversity may go unnoticed and unaddressed. Therefore, this research seeks to fill the gap by identifying academic performance patterns among PTIK students through clustering analysis. This will not only inform curriculum improvements but also support institutional decision-making for targeted academic interventions.

Informatics and Computer Engineering Education (PTIK) is a challenging program that combines technical, theoretical, and practical skills, requiring students to perform well across diverse subject areas. With the emergence of Big Data and the increased availability of educational datasets, clustering techniques can now be applied to identify distinct groups of students based on their academic performance across multiple subjects. Clustering method, especially K-Means, is one of the commonly used techniques in data mining to categorize data into groups based on similar features [8]. In particular, K-Means clustering enables a structured analysis of students' grades, providing insights into their strengths and weaknesses across different domains. Similarities between students can be found through clustering analysis of student evaluation scores using the K-means technique [9]. In a study titled "K-Means Algorithm for Grouping Student Thesis Topics," the K-Means algorithm was also used. This led to the grouping of students based on their areas of expertise. The grouping with the highest cluster indicates that the students are proficient in each group of areas of expertise, allowing them to select essay topics that are appropriate for their group of areas of expertise [10]. By clustering students according to academic performance in various categories, institutions can gain a deeper understanding of the distribution of student achievement and the specific challenges faced by certain groups.

One technique that can be used to maximize the k-means method in forming or determining the number of clusters is the silhouette coefficient [11]. To evaluate the quality of clustering, this study also utilizes the Silhouette Score, a metric that measures how well each data point is separated from others in different clusters. Although the Silhouette Scores indicate that some clusters have stronger separations than others, the clustering approach provides a foundational understanding of the varied academic achievements among PTIK students. The results offer valuable insights that can be used to tailor educational programs, such as targeted support for students in lower-performing clusters and advanced projects for high achievers. Ultimately, this research highlights the potential of data mining in identifying and addressing student needs, creating a pathway for more effective and individualized learning interventions.

Through this research, it is expected that the insights gained will inform the design of learning policies and support programs that meet the specific needs of each student group, fostering a more adaptive and supportive educational environment in the PTIK program.

II. METHOD

With an emphasis on the 2022 cohort, this study employs a quantitative methodology to examine academic performance trends among students enrolled in Sebelas Maret University's Informatics and Computer Engineering Education (PTIK) program. The main technique is data mining using K-Means clustering to find unique student groups according to their academic achievement in a variety of categories.

A. Data Collection

The dataset used in this study comprises the academic records of all students enrolled in the 2022 cohort of the Informatics and Computer Engineering Education (PTIK) program at Sebelas Maret University. This dataset includes a total of 87 students, covering academic grades from the first to fourth semesters. The data consists of average scores derived from multiple subjects that are grouped into specific course categories, such as Software Engineering (RPL), Multimedia, Education, General Courses, and Expertise-related courses. Each student's performance in these categories serves as the basis for the clustering analysis.

The academic data used in this study were obtained from the PTIK UNS Study Program administrator in the form of student transcripts. Each course was then grouped into one of six predefined categories based on the academic focus of the subject: Software Engineering (RPL), Computer Networking, Multimedia, Expertise, Education, and General Courses. The grouping was conducted by referring to the official course descriptions published on the PTIK Study Program's academic website: https://ptik.fkip.uns.ac.id/akademik/daftar-mata-kuliah/. This classification ensured that related courses were analyzed collectively within their respective domains. Grouping features based on academic content is a commonly applied practice in educational data mining to increase interpretability and analytical relevance [12].

B. Data Preprocessing

To guarantee the precision and dependability of the clustering procedure, data pretreatment was carried out. This involved classifying courses into predetermined groups, managing missing numbers, and normalising grade scales. To prepare the dataset for clustering analysis, several preprocessing steps were conducted: categorizing course types, handling missing values, and normalising grade scales. Incomplete records were excluded to preserve data integrity. This rigorous preprocessing ensures that the input is suitable for clustering algorithms like K-Means. Proper data preprocessing, such as normalization and removal of incomplete entries, is fundamental in clustering educational datasets to achieve accurate and unbiased student groupings [13]. To preserve the quality of the data, any incomplete records were not included in the study. This method focuses on transforming the dataset to ensure it is suitable for clustering algorithms and data mining tools [14]. Preprocessing steps such as normalization and missing value handling are crucial to ensure data consistency and improve clustering results [15]. Once the data is collected, a data cleaning process is performed to deal with missing or incomplete data [16].

C. Clustering Method: K-Means

The efficiency of the K-Means clustering method in dividing data points into discrete clusters according to similarity led to its selection. For the clustering procedure, the following actions were taken:

1. Finding the Ideal Number of Clusters: The Elbow Method, which examines the sum of squared distances inside clusters, was used to determine the ideal number of clusters. Four clusters were identified as the best arrangement for this dataset using this strategy.
2. Implementation of Clustering: RapidMiner, a data mining program that makes it easier to use clustering algorithms effectively, was used to carry out the clustering. In order to identify performance trends within each cluster, students were categorised according to how similar their academic performance was throughout the designated areas.
3. Interpretation of Cluster Characteristics: After generating the clusters, each was analyzed to identify its specific characteristics based on average performance in each course category. This analysis aimed to provide insights into the academic tendencies of students in each cluster.

D. Cluster Evaluation: Silhouette Score

To assess the quality of the clusters, the Silhouette Score was calculated for each cluster, measuring how similar each data point is to its assigned cluster compared to other clusters. The score ranges from -1 to 1, with higher values indicating better-defined clusters. The Silhouette coefficient is one of the most commonly used metrics to evaluate clustering quality, especially for methods like K-Means [17]. This evaluation provided a measure of how well-separated and homogenous the clusters were, with Cluster 4 exhibiting the highest score (0.458), indicating a stronger separation than the others.

III. Result And Discussion

The K-Means clustering analysis results are shown in this section along with a discussion of the unique patterns of academic achievement seen among PTIK students in the 2022 cohort. To determine the optimal number of clusters, the Elbow Method was applied by plotting the Within-Cluster Sum of Squares (WCSS) against various values of k, ranging from k=2 to k=10. The Elbow Method helps identify the point at which increasing the number of clusters yields diminishing returns in variance reduction. Based on the plotted curve, a clear "elbow" was observed at k=4, indicating that four clusters offer the best trade-off between clustering accuracy and model simplicity. This approach is widely accepted and commonly used in educational data mining to select the appropriate number of clusters [18]. Therefore, k=4 was chosen as the optimal number of clusters for further analysis. Following this, the initial centroids were randomly selected from the student dataset, and each cluster centroid represents a set of averaged scores across four categorized academic domains.

Although Clusters 1, 2, and 3 produced relatively low Silhouette Scores (ranging from 0.149 to 0.190), this does not necessarily indicate poor clustering quality, but rather reflects overlapping characteristics among some student groups. In educational datasets, especially those derived from multi-dimensional academic records, moderate to low Silhouette Scores are common due to the inherent complexity and interdependence of performance attributes. This phenomenon has been observed in studies applying fuzzy clustering to student data, where overlapping academic profiles result in lower separation metrics [19].

The decision to retain k=4 clusters was made based on the Elbow Method, which clearly indicated a significant inflection point at that value. This choice balances model simplicity and information capture. Nonetheless, to enhance cluster quality and robustness, future studies may consider alternative clustering algorithms. For instance, comparative analyses between K-Means and DBSCAN in educational settings have shown DBSCAN can better identify overlapping clusters and handle noise, often yielding higher Silhouette Scores . Additionally, Fuzzy C-Means allows soft membership assignments, which is useful for datasets where students may exhibit blended performance profiles . A systematic comparison using internal validation metrics like Davies–Bouldin or Calinski–Harabasz could further elucidate the optimal approach for clustering academic performance data.

A. Clustering Results

Based on the clustering analysis conducted in RapidMiner, the Elbow Method indicated that four clusters would provide the most meaningful insights. The clusters were analyzed for distinct performance characteristics across four academic categories: Software Engineering (RPL), Multimedia, Education, and General Courses. Although this study primarily focuses on descriptive cluster interpretation, the observed mean differences between clusters in each academic category indicate potentially significant group distinctions. These distinctions are further supported by the Silhouette Score values, which reflect intra-cluster homogeneity and inter-cluster separation, as follows:

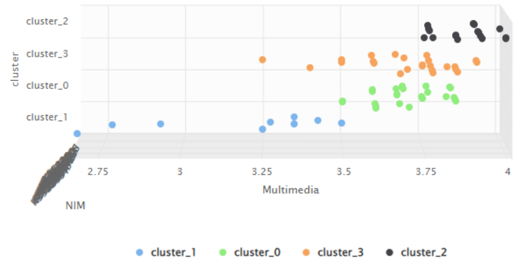

Fig. 1. Cluster 0 Visualization

Cluster 0: This cluster consists of students with moderate academic achievements across most categories, with a notable strength in Multimedia. Students in this cluster display average performance in both general and specific informatical courses. Their relatively higher performance in Multimedia suggests a potential interest or aptitude in that field, though they may benefit from additional support in other areas to achieve a more balanced skill set.

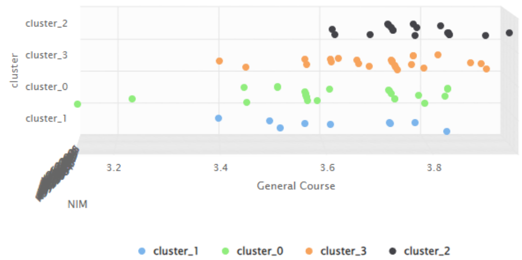

Fig. 2. Cluster 1 Visualization

Cluster 1: Students in Cluster 1 score comparatively better in General Courses but have the lowest total academic accomplishments. This pattern suggests that while students in this cluster may struggle in more complex informatics-related courses, they excel in basic or non-technical courses. To bridge performance disparities, this group can profit from focused academic assistance in informatics-related courses like computer networking and software engineering.

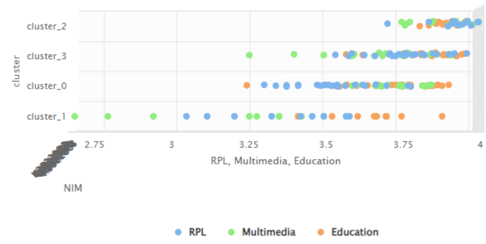

Fig. 3. Cluster 2 Visualization

Cluster 2: Students in Cluster 2 perform well in many areas, but especially in education, multimedia, and software engineering (RPL). Students in this cluster have shown a high level of academic proficiency and a broad range of skills, indicating that they can successfully understand both technical and academic topics. To help lower-performing pupils and expand their knowledge, these students can be eligible for advanced projects or peer mentoring.

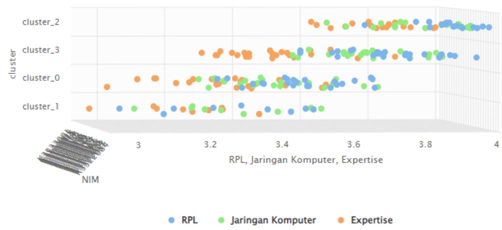

Fig. 4. Cluster 3 Visualization

Cluster 4: Cluster 3 includes students who excel in Software Engineering (RPL) and Computer Networking but have relatively lower achievements in the Expertise category. This cluster indicates a preference or strength in specific technical domains rather than a broader expertise across Informatics fields. Students in this cluster might benefit from exposure to additional resources in the Expertise category to develop a more comprehensive skill set.

B. Cluster Evaluation Using Silhouette Score

To assess the internal quality of each cluster, a Silhouette Score was calculated for each, yielding the following results:

Table I. Silhouette Score

	Silhouette scores
Cluster 0	0,149
Cluster 1	0,190
Cluster 2	0,181
Cluster 3	0,458

A moderate degree of separation across clusters is shown by the Silhouette Score values. With the highest Silhouette Score (0.458), Cluster 3 showed more uniform traits and distinct distinction between its members. Clusters 0, 1, and 2, on the other hand, received comparatively low ratings, indicating that some of the students in these clusters have traits in common with those of other clusters. This might be because academic performance in other categories is similar, which could indicate that some disciplines have overlapping learning patterns. While K-Means provides an efficient and straightforward approach, it has notable limitations in this context. First, it assumes spherical and equally sized clusters, which may not reflect the actual distribution of student academic performance that often exhibits varied shapes or densities. Second, K-Means is highly sensitive to outliers and noisy data, extreme student scores can disproportionately influence centroid positions and skew cluster assignments [20]. Third, the algorithm requires a predefined number of clusters (k), which might not always align with the natural groupings in educational datasets.

These limitations could impact the reliability of the clustering results, particularly in how borderline students or outlier performances are categorized. For instance, students with mixed academic profiles may be forced into clusters that do not accurately reflect their learning trajectories, and outliers may distort cluster centroids. Consequently, findings based solely on K-Means should be interpreted with caution. To mitigate these issues, future research could explore more robust alternatives such as DBSCAN, which can handle non-spherical clusters and identify noise; Fuzzy C-Means, which accommodates overlapping cluster memberships; or K-Medoids, which is less influenced by outliers. Comparative evaluations using internal metrics (e.g., Davies-Bouldin, Calinski-Harabasz) and outlier-aware variants of K-Means would further strengthen the validity of conclusions drawn from cluster analysis.

C. *Interpretation of Clusters and Educational Implications*

Each cluster provides insights into students' academic needs and potential areas for targeted support or enhancement, as detailed below:

- Cluster 0: The moderate achievements of students in this cluster, coupled with strength in Multimedia, suggest a need for academic support in other Informatics fields. Interventions could focus on strengthening skills in core Informatics subjects to achieve a balanced academic profile, while fostering their interest in Multimedia through specialized projects or resources.
- Cluster 1: As the cluster with the lowest academic performance, yet with relative strength in General Courses, Cluster 1 may benefit from an intensive support program in technical subjects. This might include foundational workshops, tutoring in Software Engineering and Computer Networking, or remedial courses to build a stronger foundation in Informatics-related subjects.
- Cluster 2: Students in this cluster demonstrate balanced, high performance across categories, positioning them as candidates for advanced learning opportunities. Providing mentorship roles or participation in research projects could not only further develop their skills but also enhance the learning experience for other clusters, particularly Clusters 0 and 1.

Cluster 3: This group's high performance in Software Engineering and Computer Networking suggests a focused interest in specific technical areas. Targeted support in Expertise-related courses could be valuable in expanding their academic breadth, while opportunities for deeper specialization in their areas of strength could also be beneficial.

IV. Conclusions

The clustering results from this study offer practical insights that can assist academic advisors and program managers in tailoring support strategies for students. By identifying groups of students with similar academic performance patterns, advisors can more effectively design personalized interventions. For instance, students in clusters characterized by strong performance in technical domains but lower results in general or education-related courses may benefit from academic writing workshops or soft-skills enhancement programs. Conversely, students in clusters with strong general course performance but low scores in Multimedia or RPL could be offered targeted tutoring or peer mentoring in those specific areas.

The interpretation of cluster characteristics in this study was conducted based on domain understanding of the PTIK curriculum structure and academic performance expectations across categorized subjects. Although no formal external validation such as expert review or stakeholder consultation was conducted, the course groupings and observed performance patterns were analyzed in reference to official course descriptions and academic benchmarks outlined by the PTIK study program. This approach ensures that the clustering results maintain practical relevance to real-world academic settings. Nevertheless, we acknowledge the importance of involving domain experts, such as curriculum designers or academic advisors, to validate the semantic coherence of each cluster, particularly in confirming whether groupings align with observable student learning trends. Future research could enhance the robustness of interpretation by incorporating expert validation sessions or using labeled data to conduct external cluster validation.

In order to identify discrete student groups based on academic performance across many course categories, this study used K-Means clustering to examine the academic performance patterns of PTIK students at Sebelas Maret University, specifically among the 2022 cohort. Four distinct clusters, each of which represented distinct academic tendencies and strengths, were produced by applying the Elbow Method to determine the ideal number of clusters. The clustering results provide valuable information on the academic profiles of the students, identifying areas that could use focused educational support.

Overall, this study demonstrates the potential of data-driven methods like K-Means clustering to inform academic support strategies, ensuring that student needs are addressed with precision and relevance. By leveraging insights from such clustering analyses, educational institutions can create more targeted learning environments, ultimately contributing to improved student outcomes and more effective curriculum planning.

For future improvements, it is recommended to explore hybrid approaches that combine clustering with classification methods. For example, Oyelade et al. (2010) found that integrating K-Means clustering with deterministic/statistical models significantly enhanced performance prediction accuracy—suggesting a hybrid clustering–classification pipeline could be beneficial in profiling PTIK students [21]. Additionally, alternative clustering techniques such as Fuzzy C-Means or Self-Organizing Maps (SOM) can be tested to capture non-crisp student membership and visualize multi-dimensional academic profiles more intuitively. Rffectiveness of integrating multi-dimensional feature fusion in student performance analysis, suggesting that combining temporal and spatial features can lead to more nuanced clustering insights [22]. Future studies may also incorporate statistical validity checks like silhouette analysis or ANOVA to quantitatively verify cluster separation and include longitudinal data to monitor how student cluster membership evolves over time.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Sebelas Maret University, particularly the Informatics and Computer Engineering Education Study Program (PTIK), for their support and for providing the data necessary to conduct this research.

REFERENCES

[1] Romero C, Ventura S. Educational data mining and learning analytics: An updated survey. *WIREs Data Mining Knowl Discov*. 2020; 10:e1355. https://doi.org/10.1002/widm.1355.

[2] Khan, A., Ghosh, S.K. Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Educ Inf Technol* 26, 205–240 (2021). https://doi.org/10.1007/s10639-020-10230-3.

[3] Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1). https://doi.org/10.1186/s40561-022-00192-z

[4] Simangunsong, B. N., Manalu, M. R., & Medan, U. I. (2023). Testing the K-Means Clustering Algorithm in Processing Student Assignment Grades Using the RapidMiner Application. *Journal of Data Science*, *1*, 51–60.

[5] Casquero, O., Ovelar, R., Romo, J., Benito, M., & Alberdi, M. (2016). Students' personal networks in virtual and personal learning environments: a case study in higher education using learning analytics approach. *Interactive Learning Environments*, *24*(1), 49–67. https://doi.org/10.1080/10494820.2013.817441

[6] B. Minaei-Bidgoli, D. A. Kashy, G. Kortemeyer and W. F. Punch, "Predicting student performance: an application of data mining methods with an educational Web-based system," *33rd Annual Frontiers in Education, 2003. FIE 2003.*, Westminster, CO, USA, 2003, pp. T2A-13, doi: 10.1109/FIE.2003.1263284

[7] Khan, A., Ghosh, S.K. Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Educ Inf Technol* **26**, 205–240 (2021). https://doi.org/10.1007/s10639-020-10230-3

[8] Rahayu, N. D., Anshor, A. H., & Afriantoro, I. (2024). Penerapan Data Mining untuk Pemetaan Siswa Berprestasi menggunakan Metode Clustering K-Means. JUKI : Jurnal Komputer Dan Informatika, 6(1), 71–83. https://doi.org/10.53842/juki.v6i1.474

[9] Liu, R. (2022). Data Analysis of Educational Evaluation Using K-Means Clustering Method. Computational Intelligence and Neuroscience, 2022. https://doi.org/10.1155/2022/3762431

[10] M. R. Muttaqin and M. Defriani, "Algoritma K-Means untuk Pengelompokan Topik Skripsi Mahasiswa," Ilk. J. Ilm., vol. 12, no. 2, pp. 121–129, 2020, doi: 10.33096/ilkom.v12i2.542.121-129

[11] Hartama, D., & Anjelita, M. (2022). Analysis of Silhouette Coefficient Evaluation with Euclidean Distance in the Clustering Method (Case Study: Number of Public Schools in Indonesia). Jurnal Mantik, 6(3), 3667–3677. https://iocscience.org/ejournal/index.php/mantik/article/view/3318.

[12] Mashael A. Al-Barrak and Muna Al-Razgan, "Predicting Students Final GPA Using Decision Trees: A Case Study," *International Journal of Information and*

Education Technology vol. 6, no. 7, pp. 528-533, 2016, DOI: 10.7763/IJIET.2016.V6.745

[13] et. al., M. V. . (2021). Evaluating Students Placement Performance Using Normalized K-Means Clustering Algorithm. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(11), 3785–3792. https://doi.org/10.17762/turcomat.v12i11.6488

[14] Mohamed Nafuri, A. F., Sani, N. S., Zainudin, N. F. A., Rahman, A. H. A., & Aliff, M. (2022). Clustering Analysis for Classifying Student Academic Performance in Higher Education. *Applied Sciences (Switzerland)*, *12*(19). https://doi.org/10.3390/app12199467

[15] Lhoucine Bahatti, Omar Bouattane, My Elhoussine Echhibat and Mohamed Hicham Zaggaf, "An Efficient Audio Classification Approach Based on Support Vector Machines" International Journal of Advanced Computer Science and Applications(ijacsa), 7(5), 2016. http://dx.doi.org/10.14569/IJACSA.2016.070530

[16] Al-Hagery, M. A., Alzaid, M. A., Alharbi, T. S., & Alhanaya, M. A. (2020). Data Mining Methods for Detecting the Most Significant Factors Affecting Students' Performance. *International Journal of Information Technology and Computer Science*, *12*(5), 1–13. https://doi.org/10.5815/ijitcs.2020.05.0

[17] Peter J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics, Volume 20, 1987, Pages 53-65, ISSN 0377-0427, https://doi.org/10.1016/0377-0427(87)90125-7.

[18] KETCHEN, D.J. and SHOOK, C.L. (1996), THE APPLICATION OF CLUSTER ANALYSIS IN STRATEGIC MANAGEMENT RESEARCH: AN ANALYSIS AND CRITIQUE. Strat. Mgmt. J., 17: 441-458. https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G

[19] Ovtšarenko, O. Opportunities of machine learning algorithms for education. *Discov Educ* **3**, 209 (2024). https://doi.org/10.1007/s44217-024-00313-5

[20] Iliyas Karim khan, Hanita Binti Daud, Nooraini binti Zainuddin, Rajalingam Sokkalingam, Abdussamad, Abdul Museeb, Agha Inayat. Addressing limitations of the K-means clustering algorithm: outliers, non-spherical data, and optimal cluster selection[J]. AIMS Mathematics, 2024, 9(9): 25070-25097. doi: 10.3934/math.20241222

[21] O. J. Oyelade, O. O. Oladipupo, I. C. Obagbuwa. Application of k Means Clustering algorithm for prediction of Students Academic Performance. International Journal of Computer Science and Information Security, IJCSIS, Vol. 7, No. 1, pp. 292-295. https://doi.org/10.48550/arXiv.1002.2425

[22] Luo, Z., Mai, J., Feng, C., Kong, D., Liu, J., Ding, Y., Qi, B., & Zhu, Z. (2024). A Method for Prediction and Analysis of Student Performance That Combines Multi-Dimensional Features of Time and Space. Mathematics, 12(22), 3597. https://doi.org/10.3390/math12223597