iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

COMPARATIVE STUDY OF CLUSTERING ALGORITHMS FOR STUDENT PERFORMANCE EVALUATION

Karim Md Razaul
AHUT Students, School of Computer Science and Technology, Anhui University of Technology, Ma'anshan, Anhui, China

Cheng Zekai
Associate Professor, School of Computer Science and Technology, Anhui University of Technology, Ma'anshan, Anhui, China

ABSTRACT
Predicting student performance is essential for enhancing educational outcomes, enabling educators to identify students who may need additional support or intervention. Clustering algorithms, as unsupervised data mining techniques, are particularly effective at uncovering patterns in student performance data. These algorithms can group students based on their exam scores, providing insights that allow for more tailored and targeted educational strategies. This study compares four unsupervised methods K-Means, DBSCAN, Hierarchical Clustering (Ward linkage), and Gaussian Mixture Models (GMM) on a dataset of 200 students' scores across five exam questions. After standardizing the data, we project it into two dimensions via Principal Component Analysis (PCA) for visualization. We then evaluate each model using three validation metrics: Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index. K-Means with k = 5 achieves the highest Silhouette (0.387) and Calinski-Harabasz (90.156) scores and the lowest Davies-Bouldin Index (0.883), outperforming alternatives in both visual separation and quantitative metrics. DBSCAN identifies noise but yields overlapping clusters; Hierarchical clustering shows moderate cohesion; GMM produces softer boundaries. Our results demonstrate that K-Means offers the most interpretable and robust grouping for this educational dataset, providing a practical tool for segmenting students into performance tiers. Future work may explore dynamic k-selection methods, incorporation of additional student features, and deployment in intelligent tutoring systems.

Keywords:
K-Means Clustering, Student Performance, Unsupervised Learning, Cluster Evaluation, Educational Data Mining.

1. INTRODUCTION

Evaluating student performance is a compelling problem that can be addressed effectively using data mining techniques. Beyond education, data mining plays a crucial role across various fields [1–3]. Numerous studies have analyzed student performance by leveraging historical data to identify factors influencing outcomes and predict future results [4]. This approach enables improvements in educational quality and processes [5]. Through data mining, educational institutions can discover meaningful patterns and insights [6, 7], enhancing academic performance and effectiveness [8]. Such insights are essential for addressing pedagogical challenges and developing effective teaching and learning models [9]. Ultimately, this contributes to higher levels of student satisfaction with courses and instructors [10].

The benefits of evaluating student performance include reducing educational risks such as student dropout [11–14], increasing enrollment of new students [15], and analyzing study duration [16], among others. Additionally, data mining supports increased success rates, student retention [17], and successful graduation outcomes [18, 19]. As academic institutions gather increasing volumes of student records, exam scores, and behavioral logs, the challenge shifts from data collection to extracting actionable insights. Traditional statistical techniques often fall short in revealing hidden

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

structures within the data, particularly when the goal is to identify latent groupings or patterns among students. Clustering algorithms an essential branch of unsupervised machine learning provide a powerful alternative for uncovering such latent structures without the need for labeled outputs.

Clustering refers to the task of grouping data points such that those within the same group (or cluster) are more similar to each other than to those in other groups. In the context of education, clustering can help identify students with similar learning abilities, behavioral trends, or performance profiles. For instance, by clustering students based on their responses to exam questions, one can determine which students share similar strengths or weaknesses across specific topics. These insights can then be used to customize teaching strategies, allocate tutoring resources, and monitor the effectiveness of pedagogical interventions.

In this research, we focus on clustering students based on their performance in a single course final examination. Each student is represented by a vector of five numerical scores, corresponding to five different questions in the exam. These questions cover various topics and skills within the course, making the resulting score vectors rich with diagnostic potential. Rather than relying solely on a cumulative grade or average, our approach analyzes the underlying distribution and interrelationships among these scores to form distinct performance clusters.

To achieve this, we evaluate and compare the effectiveness of four widely used clustering algorithms:

1. K-Means Clustering: A partitioning method that assigns data points to k clusters such that the within-cluster sum of squares is minimized. It is efficient and interpretable but requires the number of clusters k to be pre-specified.
2. DBSCAN (Density-Based Spatial Clustering of Applications with Noise): A density-based algorithm that groups closely packed points and identifies noise or outliers. It does not require specifying the number of clusters but can be sensitive to density parameters.
3. Hierarchical Clustering (Ward Linkage): An agglomerative approach that successively merges clusters to minimize intra-cluster variance. It creates a tree-like structure (dendrogram) and provides flexibility in choosing the final number of clusters post-analysis.
4. Gaussian Mixture Models (GMM): A probabilistic model that assumes data points are generated from a mixture of Gaussian distributions. Unlike hard-assignment algorithms like K-Means, GMM provides soft assignments based on the probability of membership in each cluster.

Each algorithm brings unique strengths and assumptions, and their performance may vary depending on the data distribution. Therefore, it is crucial to evaluate those rigorously using appropriate validation metrics.

We use a multi-step methodology: (1) data preprocessing through normalization and Principal Component Analysis (PCA) to improve comparability and enable visualization; (2) application of each clustering algorithm under controlled and consistent settings; and (3) evaluation using three standard internal validation metrics Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index. These metrics assess different aspects of cluster quality, such as compactness, separation, and dispersion.

Our primary objective is to determine which algorithm offers the most meaningful and interpretable grouping of students based on their exam scores. A secondary goal is to provide educators with a replicable clustering framework for similar performance evaluation tasks. By presenting both visual (PCA-based plots) and quantitative (metric-based) analyses, we aim to ensure both accessibility and rigor in our findings.

This research contributes to the field of Educational Data Mining (EDM) by offering a comparative framework for clustering-based performance analysis, with practical implications for curriculum planning, learning analytics, and adaptive instruction. The results not only reveal which clustering approach performs best on our dataset but also offer insights into how student score patterns can be better understood and acted upon in real-world academic settings.

2. RELATED WORK

Unsupervised learning techniques have been widely applied in educational data mining to uncover latent structures in student behavior, performance, and engagement. Early works (Jain, 2010) demonstrated the utility of K-Means for partitioning learners into achievement-based clusters, facilitating personalized feedback [20]. However, K-Means' need to predefine k and sensitivity to initialization motivated exploration of alternative methods.

DBSCAN (Ester et al., 1996) addresses these limitations by identifying dense regions without requiring k. It has proven effective in detecting outlying student behaviors (e.g., irregular assignment submissions) but can struggle when

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

cluster densities vary [21]. **Hierarchical clustering** offers a multiscale view via dendrograms, enabling educators to select clustering granularity post hoc. Rokach and Maimon (2005) applied Ward's method to group learners by forum participation patterns, highlighting its interpretability [22].

GMMs introduce probabilistic cluster assignments, accommodating uncertainty in student categorization (Rehman et al., 2022). By modeling each cluster as a Gaussian component, GMMs can capture overlapping performance tiers, although they may over fit when data deviates from normality [23].

Comparisons across these algorithms in educational contexts remain limited. Some studies combine K-Means with PCA to visualize clusters (e.g., Xie et al., 2018), while others integrate clustering with supervised prediction for dropout analysis. Yet a systematic evaluation using multiple validation metrics and consistent preprocessing has been lacking [24].

This study fills that gap by directly comparing K-Means, DBSCAN, Hierarchical (Ward), and GMM on a uniform dataset of five exam questions. By employing three cluster validity indices, we provide a robust assessment of cluster quality. Our approach aligns with best practices in clustering evaluation (Arbelaitz et al., 2013) and offers practical guidance for educational researchers seeking to segment student populations [25].

3. METHODOLOGY

3.1 Dataset Description

For this study, datasets was used to analyze student performance based on their scores across five exam questions. Our datasets were selected from the final exam of the Data Structure course at AHUT in the first semester of the 2022-2023 academic year. Datasets provide information on student responses to the final exam, with each student marked on five questions. We intentionally selected the datasets to examine how clustering reveals student performance and educational interventions. The study involves five key assessment areas:

a) Q1 - problem-solving measuring the ability to solve computational problems;
b) Q2 - theoretical understanding;
c) Q3 - algorithm analysis assessing the ability to analyze algorithms;
d) Q4 - advanced problem solving testing the ability to solve highly complex problems;
e) Q5 - case study evaluating real-world application skills.

Dataset consists of student performance on five questions in the final exam, where each question is graded based on how well it reflects the specific skills developed for that question. The following Table 4-1 is a sample representation of the data from Dataset:

Table 3-1 Sample Data of Dataset

Student ID	Q1	Q2	Q3	Q4	Q5
209074151	7	0	7	0	15
209084047	16	7	8	34	9
209084263	14	8	9	21	20
209144134	16	6	9	30	9
229074304	16	10	8	33	19
229074305	15	6	7	26	20
229074306	14	7	5	31	4
229074307	11	4	8	18	9
229074308	14	7	7	39	5

In this dataset, each row represents a unique student, and each column represents their score on one of the five questions. Specific question scores are (Q1 score range (0-20), Q2 score range (0-10), Q3 score range (0-10), Q4 score range (0-40), and Q5 score range (0-20). The performance on these questions is the primary feature used for clustering.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

The goal of analyzing this dataset is to find patterns in student responses, identify those who perform similarly, and provide targeted interventions to help struggling students or encourage high performers.
The dataset comprises N=200N students, each represented by a five-dimensional feature vector

$$\mathbf{x}_i = [s_{i1}, s_{i2}, s_{i3}, s_{i4}, s_{i5}],$$

where s_{ij} denotes the score of student i on question j. Scores range from 0 to 40.

3.2 Preprocessing

The first step involved preparing the student performance data for clustering. The procedure included cleaning the dataset by handling missing or noisy data. For this study, missing values were addressed through imputation methods, while noisy data were managed using outlier detection techniques. Then, the data was normalized using the StandardScaler to ensure that all features (questions' scores) had the same scale, which is crucial for the K-Means algorithm, as it is sensitive to the magnitude of the data. Here were the breakdown data preprocessing steps:

$$\tilde{s}_{ij} = \frac{s_{ij} - \mu_j}{\sigma_j},$$

Where μ_j and σ_j are the mean and standard deviation of question j. To facilitate 2D visualization, we apply Principal Component Analysis (PCA) to obtain

$$\mathrm{z}_i = \mathrm{W}^{\top}\tilde{\mathrm{x}}_i, \mathrm{W} \in \mathbb{R}^{5\times 2},$$

where columns of W are the top two eigenvectors of the covariance matrix of $\widetilde{\mathrm{X}}$.

3.3 Clustering Models

3.3.1 K-Means

Partition data into $k = 5$ clusters by minimizing the within-cluster sum of squares:

$$\min_{C} \sum_{m=1}^{k} \sum_{\mathbf{x}_i \in C_m} \| \mathbf{x}_i - \boldsymbol{\mu}_m \|^2,$$

where $\boldsymbol{\mu}_m$ is the centroid of cluster C_m.

3.3.2 DBSCAN

Density-based clustering group's points with ε -neighborhoods containing at least $\text{minPts} = 5$ Points not meeting this density are labeled as noise ($C = \{-1\}$).

3.3.3 Hierarchical (Ward Linkage)

Agglomerative clustering starts with each point as a singleton cluster and iteratively merges the pair (C_p, C_q) that minimizes the increase in total within-cluster variance:

$$\Delta(C_p, C_q) = \frac{|C_p||C_q|}{|C_p| + |C_q|} \| \boldsymbol{\mu}_p - \boldsymbol{\mu}_q \|^2$$

3.3.4 Gaussian Mixture Model (GMM)

Assumes data are generated from a mixture of $k = 5$ Gaussians:

$$p(\mathbf{x}_i) = \sum_{m=1}^{k} \pi_m \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_m, \Sigma_m),$$

with mixing weights $\{\pi_m\}$. Assignments are based on maximum posterior probability.

3.4 Evaluation Metrics

3.4.1 Silhouette Score

For each point i let a_i be its average intra-cluster distance and b_i the minimum average distance to any other cluster. Then

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}, s = \frac{1}{N}\sum_{i} s_i,$$

`with $s \in [-1,1]$ (higher is better).

3.4.2 Davies-Bouldin Index

Defines :

ISSN: 2456-9348
Impact Factor: 8.232

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

$$DB = \frac{1}{k}\sum_{m=1}^{k} \max_{p\neq m} \frac{S_m + S_p}{\| \boldsymbol{\mu}_m - \boldsymbol{\mu}_p \|},$$

where S_m is the average distance of points in C_m to $\boldsymbol{\mu}_m$. Lower values indicate better separation.

3.4.3 Calinski-Harabasz Index

Measures variance ratio:

$$CH = \frac{\mathrm{tr}(B_k)/(k-1)}{\mathrm{tr}(W_k)/(N-k)},$$

where B_k and W_k are between- and within-cluster dispersion matrices. Larger CH implies more distinct clustering.

4. RESULT AND DISCUSSION

4.1 Visualization of Clustering Results and Table

Figure 1 shows PCA projections colored by cluster assignments.

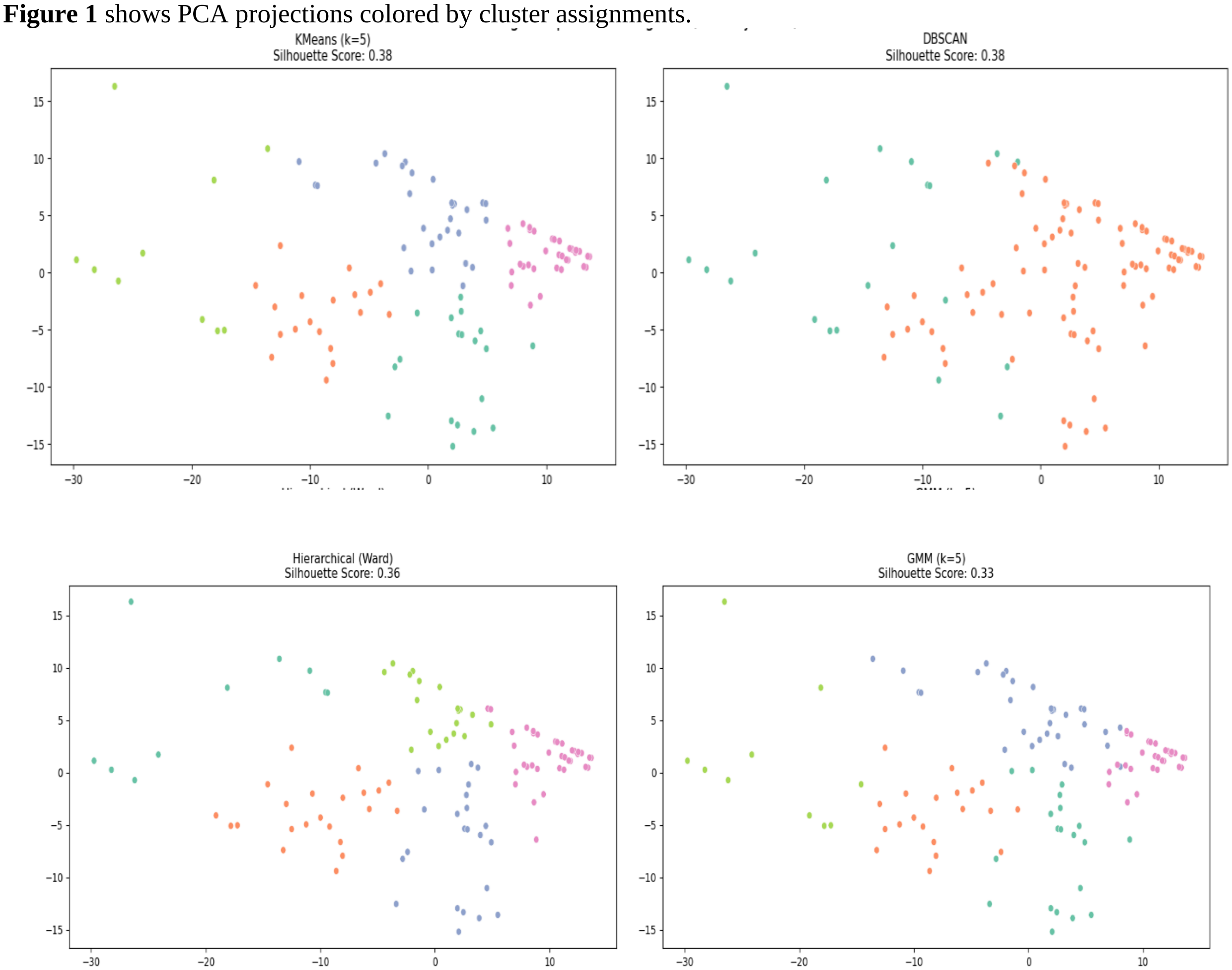

Figure 4-1(a–d). PCA scatter plots of clustering assignments

Qualitatively, K-Means (Fig. 1a) displays five well-separated ellipsoidal clusters. DBSCAN (Fig. 1b) isolates noise but merges some clusters. Hierarchical (Fig. 1c) captures structure but yields irregular shapes. GMM (Fig. 1d) produces overlapping clusters, reflecting probabilistic boundaries.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

4.2 Quantitative Comparison: Result Visualization and Table

Table 4-1 and **Figure 4-2** summarize validation metrics.

Table 4-1 Comparing Silhouette, Davies-Bouldin, and Calinski-Harabasz scores

Model	Silhouette Score	Davies-Bouldin Index	Calinski-Harabasz Index
KMeans (k=5)	0.387	0.883	90.156
DBSCAN	0.379	1.101	56.477
Hierarchical (Ward)	0.364	0.993	82.604
GMM (k=5)	0.296	1.04	69.445

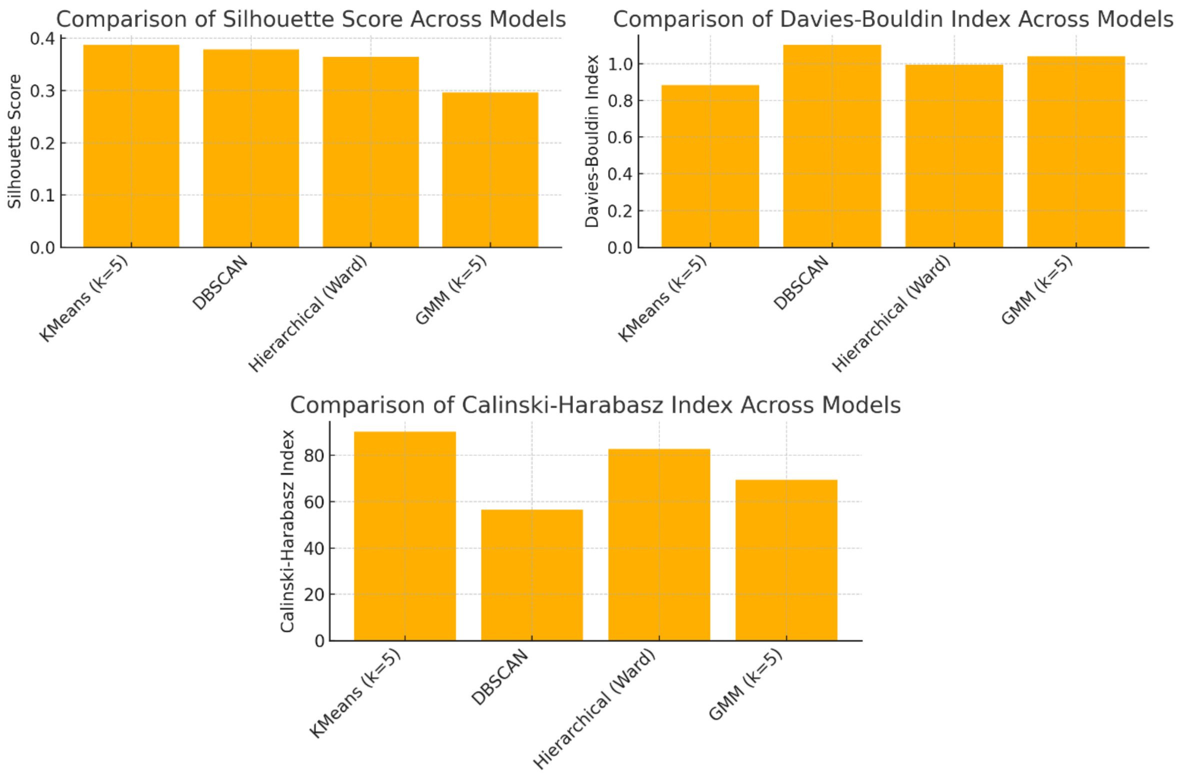

Figure 4-2 Bar charts comparing Silhouette, Davies-Bouldin, and Calinski-Harabasz scores

K-Means leads across all three metrics. Its Silhouette Score (0.387) is highest, indicating clear separation. Its Davies-Bouldin Index (0.883) is lowest, showing minimal intra- vs. inter-cluster similarity. Its Calinski-Harabasz Index (90.156) far exceeds alternatives, confirming tight, well-distributed clusters.

4.3 Result Discussion

The results clearly show that the K-Means algorithm gives the best clustering performance for our dataset. One reason for this is how K-Means works. It tries to group the students in such a way that students in the same group are very similar to each other, and students in different groups are very different. This fits well with our exam score data, where student performance has a relatively smooth and consistent distribution.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

In K-Means, each cluster is formed around a center (called a centroid), and the algorithm keeps improving these centers to make sure each student is as close as possible to their cluster's center. Since exam scores often follow a predictable pattern (some students do well, some do average, and some struggle), K-Means can easily find these patterns and group students meaningfully.
On the other hand, DBSCAN did not perform as well. This method groups students based on how close they are to each other, but it is very sensitive to a setting called epsilon (ε), which defines the neighborhood radius. If the scores are not spread evenly (for example, if there are big gaps between some students' scores), DBSCAN either creates too many clusters or merges too many students into one, making it hard to get clear groupings.
Hierarchical Clustering allows us to build a tree of student groupings and then choose how many final clusters we want. While this gives flexibility, it's more affected by unusual scores or "noise" in the data. This can cause less accurate groupings, especially when outlier students perform very differently from the rest.
Gaussian Mixture Models (GMM) take a different approach. Instead of assigning students to one group only, GMM gives each student a probability of belonging to each cluster. While this is good for showing uncertainty, it also makes the boundaries between clusters less clear. In practice, this means we don't always know exactly which student belongs to which group, which is not ideal when we want to take specific actions based on their group.
When we look at the PCA (Principal Component Analysis) visualization, we can clearly see how the K-Means clusters are separated. These plots help us understand that K-Means does a good job of grouping students in a way that makes sense visually and statistically. Each cluster appears as a tight group in the graph, with clear space between the clusters.
From an educational point of view, the K-Means clusters are easy to interpret. For example:

- One group contains students with high scores we can call this the "high achievers" group.
- Another group includes students with average scores these are the "average performers."
- The last group has students who scored low they are "students needing support."

This kind of grouping can help teachers and school administrators a lot. For example:

- Students in the high achiever group can be given extra tasks, advanced learning materials, or leadership roles in peer study sessions.
- Average performers might benefit from group work and regular feedback to help them reach the next level.
- Students needing support can be given special attention, such as tutoring, mentoring, or changes in teaching strategy.

These insights are very useful for designing teaching strategies and personalized learning plans. The clustering results make it easier to understand student performance, not just by looking at raw scores, but by seeing patterns and group behaviors.

5. CONCLUSION

This study provides a systematic comparison of four clustering algorithms K-Means, DBSCAN, Hierarchical (Ward), and GMM applied to student exam-score data. Using three standard validation metrics and PCA visualizations, K-Means with k=5 consistently outperformed alternative methods in both cluster cohesion and separation. Its highest Silhouette and Calinski-Harabasz scores, combined with the lowest Davies-Bouldin Index, confirm its suitability for segmenting learners into meaningful performance groups.
By translating cluster assignments into pedagogical cohorts, educators can design targeted interventions, allocate resources more effectively, and monitor progress across distinct student segments. While DBSCAN and Hierarchical clustering offer advantages in noise detection and multiscale analysis, respectively, their practical application requires careful parameter tuning and may yield less interpretable clusters. GMM introduces probabilistic nuance but at the cost of clear-cut group boundaries.
Future research should investigate adaptive k-selection techniques such as the "elbow" method or silhouette-based optimization and incorporate richer features (e.g., attendance, engagement metrics) to capture broader aspects of student learning. Integrating clustering into real-time learning platforms could further personalize instruction and feedback. Overall, our findings demonstrate that K-Means clustering is a reliable, interpretable, and computationally efficient approach for educational data mining, supporting data-driven decision-making in academic settings.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

REFERENCES

[1] B. R.B, Importance of Data Mining in Higher Education System,'‖ IOSR J. Humanit. Soc. Sci., 2013.

[2] S. Winiarti, H. Yuliansyah, and A. A. Purnama,Identification of Toddlers' Nutritional Status using Data Mining Approach,‖ Int. J. Adv. Comput.Sci.Appl., vol. 9, no. 1, pp. 164–169, 2018.

[3] I. Riadi, S. Winiarti, and H. Yuliansyah, Development and Evaluation of Android Based Notification System to Determine Patient ' s Medicine for Pharmaceutical Clinic,‖ 2017 4th Int. Conf. Electr. Eng. Comput. Sci.Informatics, no. September, pp. 19–21, 2017.

[4] K. Kohli and S. Birla, Data Mining on Student Database to Improve Future Performance,‖ Int. J.Comput. Appl., 2016.

[5] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider,Analyzing undergraduate students' performance using educational data mining,‖ Comput. Educ., 2017.

[6] S. K. Verma, R. S. Thakur, and S. Jaloree, Fuzzy association rule mining based model to predict students' performance,‖ Int. J. Electr. Comput. Eng.,vol. 7, no. 4, pp. 2223–2231, 2017.

[7] A. Abu, Educational Data Mining & Students'Performance Prediction,‖ Int. J. Adv. Comput. Sci.Appl., 2016.

[8] M. Anoopkumar and A. M. J. Md Zubair Rahman, A Review on Data Mining techniques and factors used in Educational Data Mining to predict student amelioration,‖ in Proceedings of 2016 International Conference on Data Mining and Advanced Computing, SAPIENCE 2016, 2016.

[9] A. M. De Morais, J. M. F. R. Araújo, and E. B. Costa,Monitoring student performance using data clustering and predictive modelling,‖ in Proceedings -Frontiers in Education Conference, FIE, 2015.

[10] J. L. Hung, Y. C. Hsu, and K. Rice, Integrating data mining in program evaluation of K-12 online education,‖ Educ. Technol. Soc., 2012.

[11] P. Paruechanon and W. Sriurai, Applying association rule to risk analysis for student-dropout in Information Technology Program,‖ J. Sci. Sci. Educ., vol. 1, no. 2, pp. 123–133, 2018.

[12] B. Pérez, C. Castellanos, and D. Correal, Predicting student drop-out rates using data mining techniques: A case study,‖ in Communications in Computer and Information Science, 2018.

[13] B. Perez, C. Castellanos, and D. Correal, Applying Data Mining Techniques to Predict Student Dropout: A Case Study,‖ in 2018 IEEE 1st Colombian Conference on Applications in Computational Intelligence, ColCACI 2018 - Proceedings, 2018.

[14] B. Mahatthanachai, H. Ninsonti, and N. Tantranont, A Study of Factors Influency Student Dropout Rate Using Data Mining,‖ Golden Teak Humanit. Soc. Sci.J., vol. 22, no. 4, pp. 46–55, 2016.

[15] N. Chaimongkol and P. Meesad, Association Rule Mining for Specific New Course,‖ J. Inf. Sci. Technol.,vol. 1, no. 1, pp. 15–22, Jun. 2010.

[16] P. Belsis, I. Chalaris, M. Chalaris, C. Skourlas, and A. Tsolakidis, The Analysis of the Length of Studies in Higher Education based on Clustering and the Extraction of Association Rules,‖ Procedia - Soc.Behav. Sci., vol. 147, pp. 567–575, 2014.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

[17] T. Mishra, D. Kumar, and S. Gupta, Mining students‘data for prediction performance,‖ in International Conference on Advanced Computing and Communication Technologies, ACCT, 2014.

[18] E. N. Ogor, Student academic performance monitoring and evaluation using data mining techniques,‖ in Electronics, Robotics and Automotive Mechanics Conference, CERMA 2007 - Proceedings, 2007.

[19] R. Bansal, A. Mishra, and S. N. Singh, Mining of educational data for analysing students‘ overall performance,‖ in Proceedings of the 7th International Conference Confluence 2017 on Cloud Computing, Data Science and Engineering, 2017.

[20] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

[21] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (pp. 226–231)

[22] Rokach, L., & Maimon, O. (2005). Clustering methods. In Data Mining and Knowledge Discovery Handbook (pp. 321–352). Springer. https://doi.org/10.1007/0-387-25465-X_15

[23] [23] Rehman, A., Imran, M., & Amin, R. (2022). A probabilistic clustering model for analyzing student learning behaviors in online platforms. Education and Information Technologies, 27, 4987–5005.

[24] Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2018). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. Computers & Education, 140, 103599. https://doi.org/10.1016/j.compedu.2019.103599

[25] Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. Pattern Recognition, 46(1), 243–256. https://doi.org/10.1016/j.patcog.2012.07.021