Volume 11, Issue 7, July – 2026 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul298

Enhancing IT Training Platforms: Adaptive Learning Strategies via K-Means and the ACEM Framework

Henoc Tenda[1*]; Ange Ntokusi[2]; Innocent Tshiamanda[3]; Celestin Kasela[4]; James Kiba[5]; Azaria Kusakana[6]; Blanchard Kangulumba[7]; Jerome Mwandoki[8]; Rodriguez Luzolo[9]; Dieumerci Kinguangu[10]

[1,2,3,4,5,6,8,9,10]Teaching Assistant, Department of English and Business Computing, Faculty of Economics, University of Kinshasa
[7]Teaching Assistant, Department of Information Technology, University of Kwango

Corresponding Author: Henoc Tenda[*]

Publication Date: 2026/07/22

Abstract: This study employed K-means clustering on behavioral data gathered from a programming-oriented IT training platform. This study aimed to characterize different learner profiles and map each profile with customized teaching strategies through the Adaptive Cognitive Enhancement Model (ACEM). The dataset includes the behavioral data of 3384 learners over a time span of six months, including activity logs, submissions, and progress of learning missions. We categorized these learners into four levels: passive, basic, intermediate, and advanced learners. We used internal metrics (Silhouette Coefficient: 0.682; Davies–Bouldin Index: 0.777; Calinski–Harabasz Index: 4349.79) to validate the Clustering, and PCA explained 98.90% of the variance in the data. These profiles supported the development of tailored teaching activities according to the ACEM plan, such as staged orientation (for lower-interest learners) to self-staged and peer-supported modules (advanced learners). This approach addresses the gap between unsupervised data exploration and pedagogically driven adaptive learning environments. The limitations of this study include the use of a single data platform and the exclusion of cognitive and emotional factors from the analysis. In future research, we plan to refine the model by applying multiple platforms and incorporating psychological data, thereby extending the concept of learner profiling to include psychological data. These findings have important implications for scalable and adaptive educational technologies.

Keywords*: K-Means Clustering, Learner Classification, Adaptive Learning, Learning Analytics, IT Training Platforms.*

How to Cite: Henoc Tenda; Ange Ntokusi; Innocent Tshiamanda; Celestin Kasela; James Kiba; Azaria Kusakana; Blanchard Kangulumba; Jerome Mwandoki; Rodriguez Luzolo; Dieumerci Kinguangu (2026) Enhancing IT Training Platforms: Adaptive Learning Strategies via K-Means and the ACEM Framework. *International Journal of Innovative Science and Research Technology,* 11(7), 646-659. https://doi.org/10.38124/ijisrt/26jul298

I. INTRODUCTION

The rapid expansion of online learning platforms has generated unprecedented opportunities for data-driven educational personalization (Khor, 2021). K-means clustering has emerged as a prominent technique for learner segmentation based on engagement and performance patterns (Jiménez et al., 2024). However, a significant gap exists between clustering analyses and practical pedagogical implementation. Most studies focus on algorithmic precision while neglecting the translation of clustering results into actionable, effective educational strategies. Traditional K-means clustering faces inherent limitations, including the requirement for predetermined cluster numbers and the assumption of spherical, uniformly distributed clusters that rarely align with complex educational datasets (Baharuddin, 2022). More critically, existing research fails to bridge the gap between learner segmentation and concrete instructional design, limiting educators' ability to implement data-driven personalization effectively.

This study addresses these limitations by integrating K-means clustering with the Adaptive Cognitive Enhancement Model (ACEM), an evidence-based instructional framework grounded in cognitive load theory and motivational scaffolding (Moubayed et al. 2020). Our approach transforms behavioral learner clusters into specific pedagogical interventions, creating personalized adaptive

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul298

learning paths based on real-time analytics. This study aims to bridge the research-practice gap in educational data mining by pursuing three primary objectives: first, to validate clustering effectiveness by demonstrating the utility of platform interaction data for identifying distinct learner profiles using established clustering quality measures; second, to establish cluster interpretability by employing Principal Component Analysis (PCA) to visualize and validate cluster separation and membership; and third, to enable pedagogical translation by integrating clustering outcomes within the ACEM framework to construct scalable, personalized learning paths aligned with identified learner profiles. This study explores two key research questions.

- How effectively can K-means clustering categorize learners into distinct profiles within an IT training platform?
- Which pedagogical strategies (scaffolding, mastery-based learning, gamified onboarding, peer mentoring) best align with each identified learner group?

This study makes important contributions to educational technology and learning analytics by bridging the gap between learner data and teaching practices through a systematic method for converting clustering results into practical instructional strategies. This study presents a comprehensive framework that combines algorithmic learner segmentation with evidence-based instructional design principles. Our findings revealed four distinct learner profiles: passive, basic, intermediate, and advanced, each requiring tailored pedagogical approaches.

These include scaffolded tutorials and gamified onboarding for novices, mastery-based projects for intermediate learners, collaborative and challenge-based activities for engaged students, and self-paced modules with peer mentoring for advanced learners. The integration of these strategies within a simplified ACEM implementation enables real-time system adaptation based on learner profiles, moving beyond traditional one-size-fits-all approaches toward truly personalized learning experiences. This study lays the foundation for developing scalable, intelligent learning environments that can enhance educational outcomes for both academic institutions and corporate training programs by improving personalization, engagement, and learner support.

II. RELATED WORK

K-means clustering has gained popularity in educational data mining. Many studies have used K-means to model students, assess their learning performance, and enhance adaptive learning systems. This review categorizes the literature into five major categories: (1) clustering toward learner profiling, (2) clustering toward academic performance, (3) clustering on engagement and behavioral data, (4) hybrid clustering, and (5) clustering in personalized and adaptive learning.

- *Learner Profiling via Clustering Techniques*

K-means clustering is well known for its speed and interpretability (Jain, 2010). It is successful in creating meaningful subgroups from various groups of learners. This approach utilizes academic factors (e.g., GPA and test scores) and engagement activities (e.g., logins) to identify patterns that are not typically revealed in a complete dataset (Nguyen & Do, 2021). K-means requires the number of clusters to be specified in advance, which can result in "arbitrary" groupings given limited domain knowledge of the data. It also assumes that clusters are spherical, evenly sized, and evenly spaced, which is seldom true for complex learner datasets (Xu and Tian, 2015).

Hybrid algorithms that integrate K-means with other algorithms, such as SVM and density-based clustering, can achieve encouraging performance in boundary detection and increased robustness (Ester et al. 1996). However, there remains a serious concern regarding how to apply clustering results to teaching practices. Although Clustering identifies learner groups, few studies have advanced beyond descriptive analysis to offer personalized instructional guidance addressing motivation, cognitive load, or emotional states. (Papamitsiou & Economides, 2014)

- *Use of Behavioral Data in Learning Analytics*

Behavioral indicators, such as time on task, submission behaviors, and activity logs, help to understand student engagement and persistence (Ifenthaler & Yau, 2020). Recent studies have highlighted the importance of using multidimensional behavioral data to enhance learner classification accuracy (Wei et al., 2023). Recent studies have highlighted the relevance of combining soft skills and lifestyle factors in clustering models for education. Despite the growing recognition of the relevance of behavioral data, its integration with cognitive and affective aspects remains largely unexamined. In contrast, previous studies typically explore behavioral or performance indicators separately, often overlooking essential factors such as motivation and emotional or cognitive load, which are crucial for adaptive learning.

- *Adaptive Learning Strategies Informed by Clustering*

Adaptive learning systems aim to personalize learning based on learners' characteristics using data-driven methods. Clustering is employed in many systems for content difficulty adaptation, establishing learning paths, and facilitating peer cooperation (Kurniawan & Hermawan, 2023). The Adaptive Cognitive Enhancement Model (ACEM) expands on these approaches by associating learner clusters with cognitive load theory and motivational scaffolding. This strategy permits the potential for just-in-time pedagogical decisions driven by data (Vygotsky, 1978; Sweller, 1988; Luo, 2024). Few researchers have fully applied cluster-based teaching in large-scale educational settings. This gap involves insufficient validation of systems such as ACEM and a lack of integration of real-time learner results in adaptive teaching strategies.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul298

➢ *Clustering for Academic Performance*

Several previous studies have employed K-means clustering to categorize students based on their academic characteristics, including their test scores, GPA, and course completion. For example, Vankayalapati et al. (2021) categorized students into low-, medium-, and high-grade groups, as well as low-, medium-, and high-course time groups, based on their test scores and course time. Brempong (2021) adopted a similar approach, utilizing high school students' scores on quizzes and tests to predict their likelihood of passing the national exam as a tool to identify those at risk of remediation.

Oyelade et al. (2010) applied a performance-based method, using students' GPAs as the standard for cluster membership, to explore their placement in achievement trajectory types. In this way, institutions could assess whether their accommodations were adequate and how well they met student needs, as well as the division of support among different accommodations and contributions to a course that traditional letter grades cannot capture. Henderi et al. (2020) further developed this approach by incorporating a temporal dimension into performance clustering for monitoring academic performance across semesters.

➢ *Clustering Based on Engagement and Behavioral Data*

In addition to performance indicators, there has been a growing awareness of the need to profile learners in terms of their behavior and engagement with systems. Erkan et al. (2024) utilized a 39-dimensional system overview axis, specifically Axis 11, to differentiate seven types of learners through the truncated log-likelihood of a given parameter vector. They found that a spectrum of interaction formats was preferred at various grade levels, with younger elementary-level children preferring game-based content and middle school students preferring diagnostic content.

Abdo et al. (2021) distinguished students by login frequency and assignment submission behaviors, promoting the customization of teacher-student interaction strategies. Talib et al. (2023) developed a multifaceted profiling approach that integrates both soft skills, such as resilience and leadership, and academic achievement. Based on this, they developed customized learning plans that exceeded conventional cognitive profiling methods. Similarly, Chang et al. (2020) incorporated lifestyle and study habits, and their clustering strategy demonstrated that complementary behavioral attributes could lead to a more explicit and interpretable clustering system.

➢ *Hybrid Models: K-means Combined with Other Techniques*

Hybrid models, such as the combination of AE (autoencoder) and k-means or the composition of k-means with other ML algorithms, are successful in addressing some of the limitations of k-means. These methods enhance the accuracy and flexibility of classification by leveraging the strengths of multiple algorithms. For example, K-Means and Support Vector Machines (SVM) collectively reinforce boundary detections among clusters and improve learning in non-spherical patterns observed in the learner data (Chatterjee & Choudhury, 2025).

Chang et al. (2020) presented a combination of K-means and density-peak clustering, which does not rely on users to set the number of clusters, with good scalability and accuracy on large-scale datasets. Early classification of learners, including the cold-start problem, has become increasingly important in adaptive educational systems. Hybrid approaches, such as combining K-means clustering with majority voting systems, have achieved promising results, with prediction accuracies of up to 75.47% (Mohd Talib et al., 2023).

These methods represent a broader trend toward more resilient and context-aware clustering techniques designed to handle the growing complexity and diversity of the educational data. Additionally, improvements in centroid initialization algorithms and enhanced models, such as Kmeans9+, have shown significant performance gains in clustering tasks, further strengthening the reliability of data-driven learner profiling (Abdulnassar & Nair, 2023).

➢ *Clustering in Adaptive and Personalized Learning*

K-means Clustering has been widely used in adaptive learning systems to support content delivery and fit learners' educational paths. Kurniawan and Hermawan (2023) applied the K-means and K-medoids algorithms to cluster learners based on their interests in different topics and attitudes towards their careers. K-means demonstrated a slightly lower silhouette coefficient of 0.786 compared to the K-medoids algorithm, which was also more suitable for course matching and personalized career guidance. Fadieieva (2023) presented an adaptive learning system that utilizes K-means to dynamically adjust the difficulty of modules through fuzzy logic and recommendation algorithms as learners progress along the learning path.

This training method resulted in a 35% decrease in skill gaps; therefore, K-means Clustering was used to cluster learning styles and user engagement indicators in adaptive learning platforms to augment personalized systems. Agha et al. (2023) used global course category clustering methods to assist institutions in identifying their knowledge-level taxonomies and reconceptualizing curricular content for low-performing groups of students. De Marsico et al. (2013) integrated the Felder-Silverman learning styles model into K-means for personalized e-learning materials in collaborative learning settings. This approach groups students according to their learning preferences and beliefs in their abilities.

➢ *Clustering Quality Measurement*

The validity of the clusters is a crucial factor in determining the robustness and usefulness of student segments. Various measures, such as the silhouette coefficient, Davies-Bouldin index, and cluster homogeneity, are commonly used to evaluate the performance of clustering (Oyelade et al., 2010; Henderi et al., 2020). These indices evaluate cluster cohesion and separation, which are essential for identifying meaningful groups of learners.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul298

Chang et al. (2020) suggested that density-peak clustering might be more stable and effective in complicated behavioral data than classical K-means when appropriate. However, there is no single criterion that provides the quality of Clustering, and this encourages researchers to use more than one validation criterion to conduct a comprehensive evaluation process.

➢ *Adaptive Cognitive Enhancement Model (ACEM)*

The Adaptive Cognitive Enhancement Model is a generic model for an e-learning decision support system that dynamically pulls data from the learning management system (LMS) regarding learners to influence or direct adaptive instructional and learning strategies. Embedded in the platform's structure, ACEM connects engagement patterns to targeted interventions that personalize data-driven learning environments (Luo, 2024).

Through these capabilities, the system is constructed from custom-made pieces, aggregating raw behavioral data into refined engagement monitors, grouping students using K-means clustering, and matching cluster profiles to individual pedagogical techniques. Extensive mining of these data enables teachers to gain just-in-time insights, personalizing interventions to address the needs of all students with effectiveness and precision. ACEM integrates learning design with teaching to the test, enabling teaching flexibility and the creation of learner-centered digital education environments through a combination of cluster analysis and applied teaching methods. (Sharisha et al., 2025)

➢ *Research Gap and Contribution*

Although K-means clustering is a popular clustering algorithm in the field of educational data mining, few studies have been conducted on academic success and engagement as separate features. Most studies do not merge behavioral results into the educational structure; only a few relate clustering results to new educational projects or strategies for personalizing learning systems. While some studies have investigated cluster-based learner segmentation, few empirical studies have applied such segments to examine meaningful cluster-focused instructional approaches in IT training contexts.

The present study fills this gap by providing empirical evidence for the existence of learner profiles and showing that different learning strategy conditions can be implemented with different subgroups depending on their behavioral characteristics. This bridges the gap between unsupervised learner categorization and tangible course design for adaptive learning.

III. METHODOLOGY

This study proposes a quantitative, data-driven method for categorizing learners of an IT training platform based on their behavior and accomplishments. Using unsupervised machine learning, specifically K-means clustering, this study uncovers the latent patterns in learner trace data. These insights contribute to the development of personalized learning methods.

This study adheres to the Learning Analytics Framework (LAF), which encompasses data collection, analysis, interpretation, and feedback to facilitate personalized learning. Through this approach, the system integrates learners' profiles with adaptive learning strategies to enhance personalized learning experiences within digital environments.

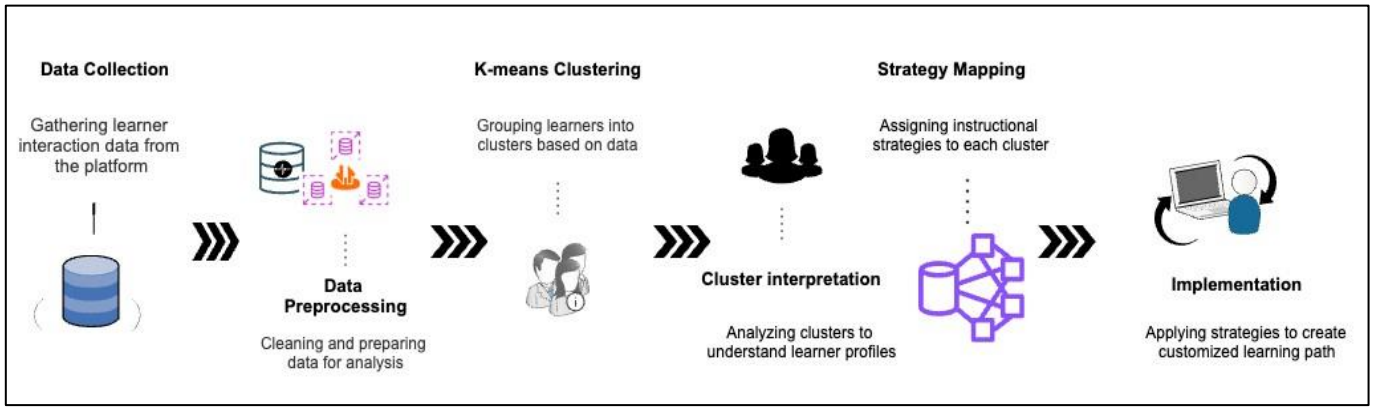

Fig 1 Cluster-Based Adaptive Learning Workflow

Figure 1: The above workflow shows the process of building customized learning methods based on learner data analysis on an IT training platform. It begins with tracking learner interactions, activity logs, task submissions, lesson completions, and other relevant data. This led us to prepare the data. Learners were clustered into various engagement profiles using K-means clustering and then compared and classified into different engagement levels (e.g., passive, basic, intermediate, and advanced). These profiles, characterized by distinct strengths and weaknesses, inform instructional strategies within the Adaptive Cognitive Enhancement Model (ACEM)

➢ *Data Description and Preprocessing*

This study uses anonymized behavioral data from 3,384 learners on a programming IT training platform over six months (January–June 2024). The dataset includes important interaction parameters, such as the total lessons attempted, lessons completed, assignment submission frequency, and earned experience points (XP). These

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul298

elements indicate the learner's engagement and performance in the learning process. The dataset is not personally identifiable, thus maintaining the learner's privacy and confidentiality. The data were initially imported into a Pandas DataFrame for the initial discovery analyses. This involves the type, structure, and handling of missing values in the data.

Several key numeric fields, such as total lesson experience ('total_lessonExp'), average micro-test score ('avg_microTestScore'), and total submitted count ('total_submittedCount'), contain missing values. To fill these gaps, we used the median of each column for imputation because it is robust against outliers and preserves the central tendency of the data. Records missing essential timestamp fields 'first_activity' and 'last_activity' were excluded, as a well-defined core timestamp element is necessary for an accurate behavioral analysis. Using the interquartile range (IQR) method, we identified outliers in features such as the total lessons attempted, lesson experience, submissions, and completed lessons.

Values outside this range were capped at their respective maximum or minimum values. This procedure effectively downplays the effect of gross outliers through light trimming of the internal structure of the dataset. We also rounded the integer-typed features after rounding, as our objective was to maintain consistent data types. Then, we normalized all features using min-max scaling to the range of [0, 1]. This normalization is enforced to weight the features equally in the clustering distance calculations, which increases the reliability and interpretability of the generated segments.

➢ *Comparative Clustering Analysis*

To support our choice of K-means clustering, we conducted a comparative analysis using different clustering methods, including hierarchical Clustering and DBSCAN. Hierarchical Clustering produces useful dendrograms that highlight the relationships among learners; however, its poor computational performance restricts its general applicability to large datasets. DBSCAN efficiently identified the core clusters and handled noise but struggled with varying density levels and was less interpretable for teaching interventions. K-means Clustering, on the other hand, provided better scaling and visualization, aligning more closely with our teaching objectives. The validity of K-means clustering was supported by a high Silhouette Coefficient, effectiveness of the Davies–Bouldin Index, and soundness of the Calinski–Harabasz scores. Hence, K-means was found to be the most suitable for efficient and actionable classification of learners.

- *Exploratory Data Visualization of Selected Learner Features*

To gain an initial understanding of learner activity patterns, we conducted exploratory data analysis (EDA) using histograms and scatter plots of the four primary features selected for clustering: total_lessonExp (total experience), total_submittedCount (assignment submission frequency), total_completed_lessons (number of lessons completed), and total_lessons_attempted (number of lessons attempted). These features were chosen because they measure the essential elements of learner engagement and performance on the platform. The analysis of the relationship between these features demonstrated that they add complementary information and can be used to achieve effective and fine-grained segmentation of learners.

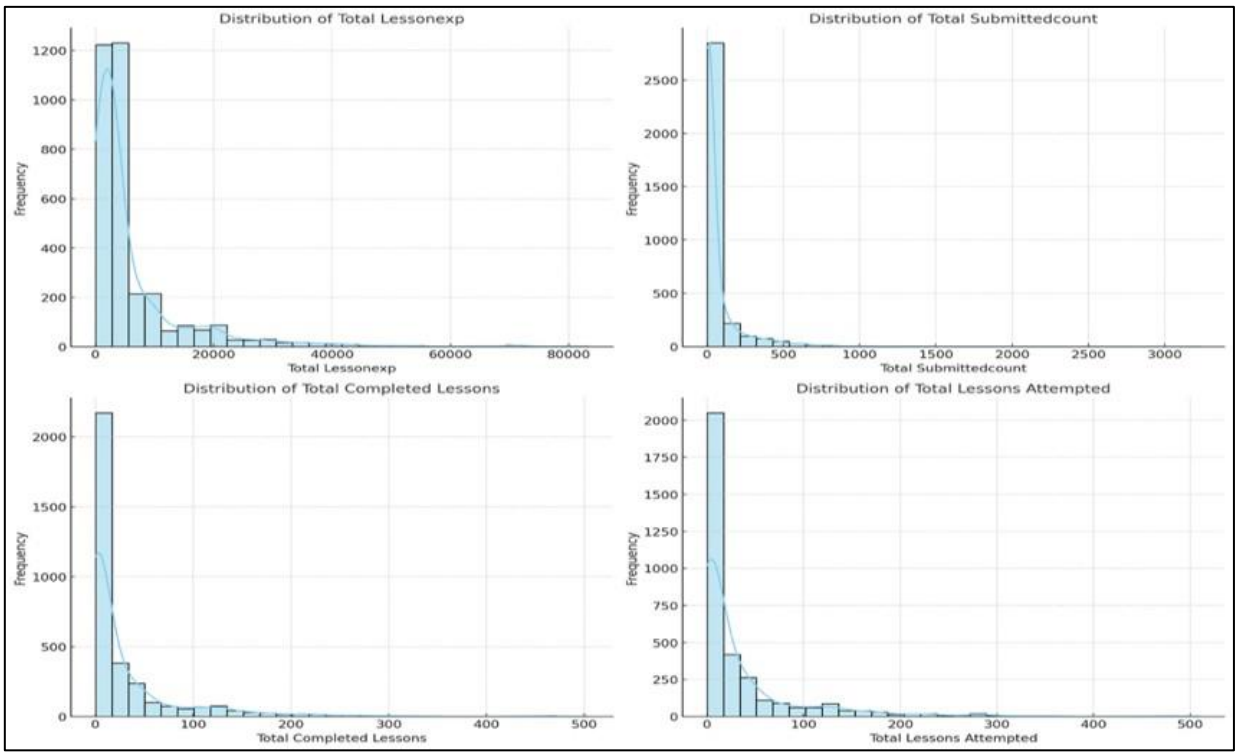

Fig 2 Data Visualization of Learner-Selected Features

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul298

The histograms of these four variables are presented in Figure 2, which are all noticeably right skewed. Most learners displayed low engagement (fewer completed courses and fewer submission counts), whereas a minority exhibited high engagement.

- *Clustering Rationale and Procedure*

We chose K-means clustering as a representative technique to segment massive numerical behavioral data because it is effective, efficient, and interpretable. This unsupervised learning approach identifies latent learner groups that perform clicks and read online, even when no prior data are available on them. We also set the number of clusters (k = 4) using the elbow method (a method of finding the point of diminishing returns from minimizing within-cluster variance) and domain expertise. This choice compromises model complexity and teachability, allowing us to divide the learners into four profiles (passive, basic, intermediate, and advanced). To ensure the quality of clustering, we used several internal validation indices.

✓ *Silhouette Coefficient:*

Measures how closely the objects in one cluster are compared to the objects in their nearest clusters.

✓ *Davies–Bouldin Index:*

This is used to compare the similarity between clusters, with a smaller index value indicating a better clustering performance.

✓ *Calinski–Harabasz Index:*

Evaluates the compactness and separation of clusters, with higher values being better.

PCA was also used to facilitate dimension reduction and investigate cluster separability, demonstrating the feasibility of our clustering method.

➢ *Using the Confusion Matrix for Cluster Validation*

To evaluate the performance of K-means clustering in practice, we computed a confusion matrix comparing the learner profiles annotated by human experts with those generated by the clustering algorithm. The confusion matrix shows that clustering has a good overlap with expert labeling for passive and advanced learners. Clustering captured a good mix of engaging and nonengaging learners. However, there is a confusing term that does not follow an established binary classification system. This suggests overlapping behaviors between the clusters, warranting further analysis to clarify these distinctions. Overall, our methodology validates the use of K-means clustering for actionable learner segmentation in adaptive instructional design.

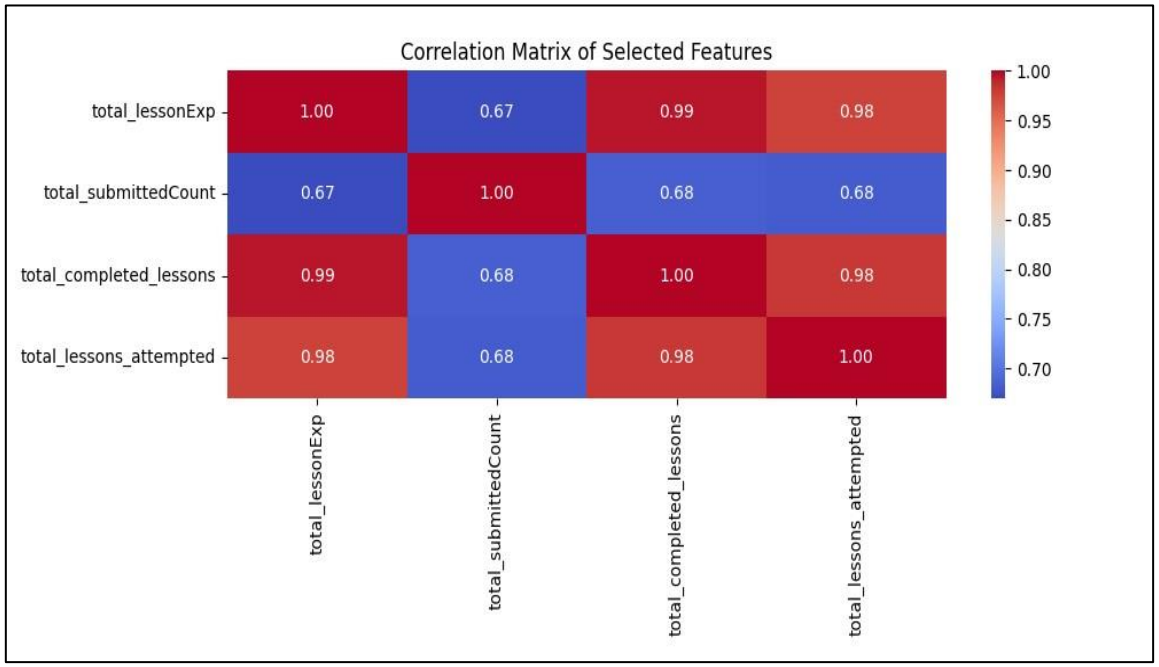

Fig 3 Confusion Matrix of Selected Features

As depicted in Figure 3, the matrix captures the similarity between the clusters derived by the algorithm and the known learner categories. It provides an overview of the classification performance and identifies ranges where overlaps between learners or uncertainties about learner segmentation may occur.

➢ *Determining Optimal Clusters*

The selection of an appropriate number of clusters is crucial for obtaining good segmentation results. Domain knowledge (e.g., novice, intermediate, expert) can inform decisions; however, objective and data-driven assessments are necessary to minimize the bias. This can be achieved using the elbow method. It graphs the Sum of Squared

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul298

Errors (SSE) for a range of clustering amounts (k). The "elbow" point, where the reduction of SSE tapers off, is a good compromise between the accuracy and complexity of the model. In our study, the elbow method indicated that three or four clusters may be reasonable choices. We selected four groups to characterize the diversity of the patterns of student engagement observed in the data. This is consistent with educational categories that divide learners into passive, basic, intermediate, and advanced levels. This choice enhances the model's interpretability, allowing for the derivation of practical suggestions for educational intervention.

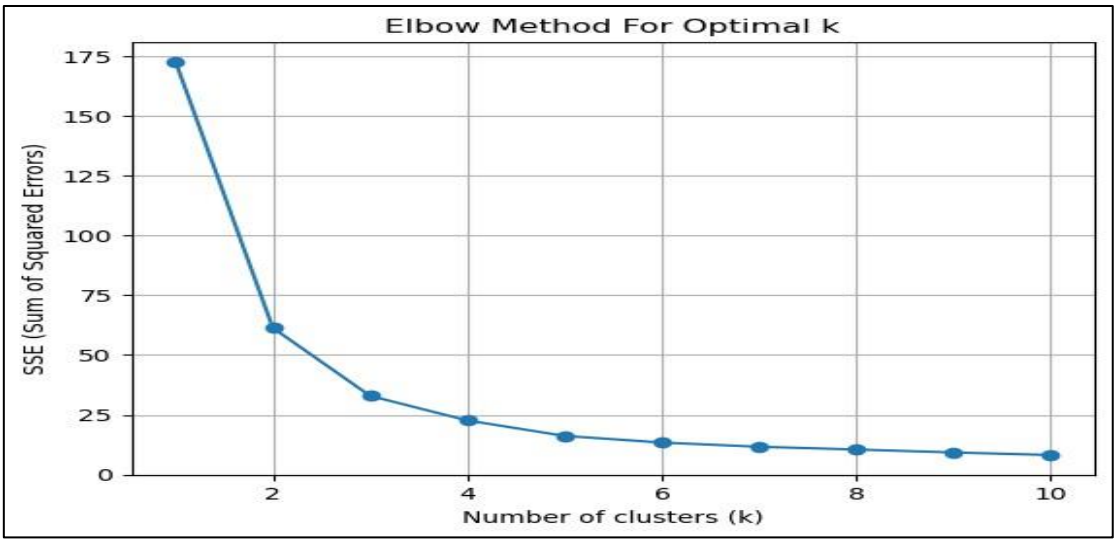

Fig 4 The Elbow Method Graph

Figure 4 illustrates the results of the elbow method for finding the ideal number of clusters for K-means clustering. The graph illustrates the relationship between the number of clusters (k) and the Sum of Squared Errors (SSE), which measures the amount of variation within the clusters. As expected, the SSE decreased as **k** increased, indicating tighter clusters. However, after k = 3 or 4, the decrease in SSE significantly slows down, marking this point as the "elbow" in the literature. This suggests an appropriate balance between the effectiveness of clustering and model complexity. This supports our conclusion that three or four clusters from the dataset represent the most suitable number of clusters.

To further evaluate the correctness of the clusters, we computed the average silhouettes for the number of clusters between two and ten (Figure 5). The silhouette score measures the similarity of an object to its cluster (cohesion) compared to other clusters (separation), where high values indicate that the object is well matched to its cluster. Our analytical method yielded the highest silhouette score at k = 2, indicating that two clusters represented the statistically optimal separation. However, the k = 4 clusters were selected based on their better interpretability of students' behavior in relation to our educational goals, despite the slightly lower silhouette score. This trade-off enables a more precise description of learner profiles while also achieving reasonable cluster validity.

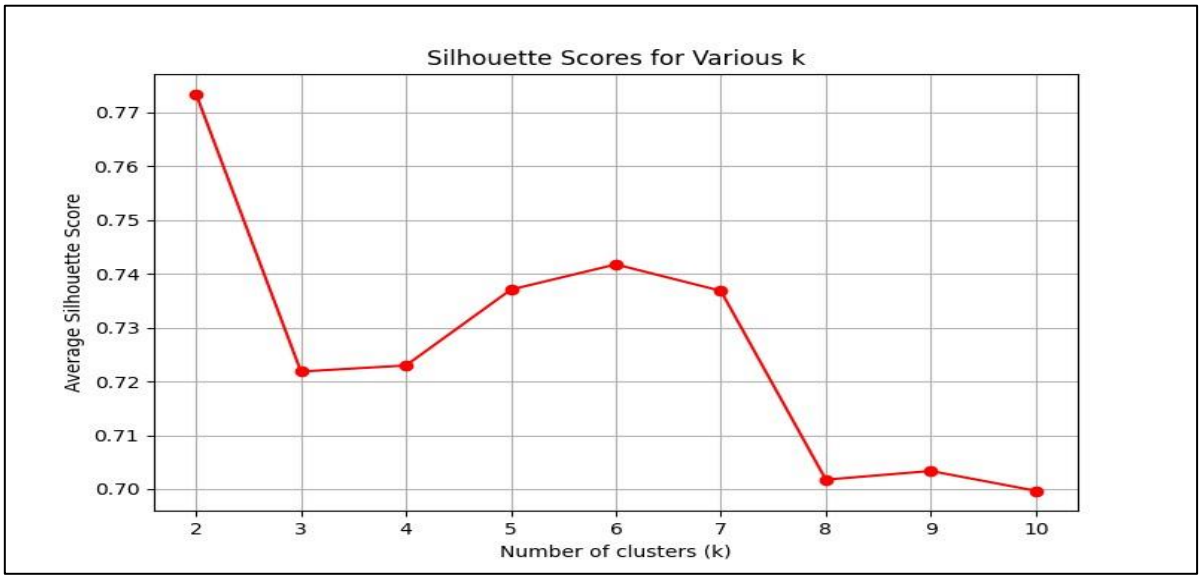

Fig 5 Silhouette Scores for K-means Clustering at Different Cluster Counts (k)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul298

Figure above (5) displays the average silhouette scores for K-means clustering on learner data with cluster counts ranging from 2 to 10. The maximum silhouette score for k = 2 suggests the best separation of the clusters. However, we selected additional clusters (k = 4) for a deeper and more useful interpretation of the learner profiles.

➢ *Clustering, Labeling, Auto-Level, and Evaluation Metrics*

Next, we calculated the average lesson experience for each cluster formed using the K-means algorithm. This served as an engagement indicator. We then organized the clusters by increasing the average total lesson experience and assigned auto-levels ranging from 0 to 3. This classification allows us to analyze learner profiles based on engagement levels. We used three standard evaluation metrics to assess the internal validity and quality of clustering results. These metrics indicate the independence and cohesion of the clusters. The findings are summarized as follows.

Table 1 Clustering Evaluation Metrics and Interpretation

Metric	Value	Interpretation
Silhouette Coefficient	0.682	High: strong cohesion within clusters and good separation between them
Calinski–Harabasz Index	4349.79	High: indicates compact, well-separated clusters
Davies–Bouldin Index	0.777	Low: minimal overlap between clusters

The Silhouette Coefficient of 0.682 shows strong cohesion within groups and effective separation from other groups. The Calinski–Harabasz Index of 4349.79 indicates high quality with distinct clusters. The Davies–Bouldin Index of 0.777 demonstrates a limited overlap between the clusters. Together, these indices highlight the strength and effectiveness of the clustering algorithm in distinguishing significant learner segments in the learning process. To further illustrate how learners are distributed across the identified groups, Figure 6 presents a scatter plot of the clusters based on the total number of lessons attempted and total lesson experience.

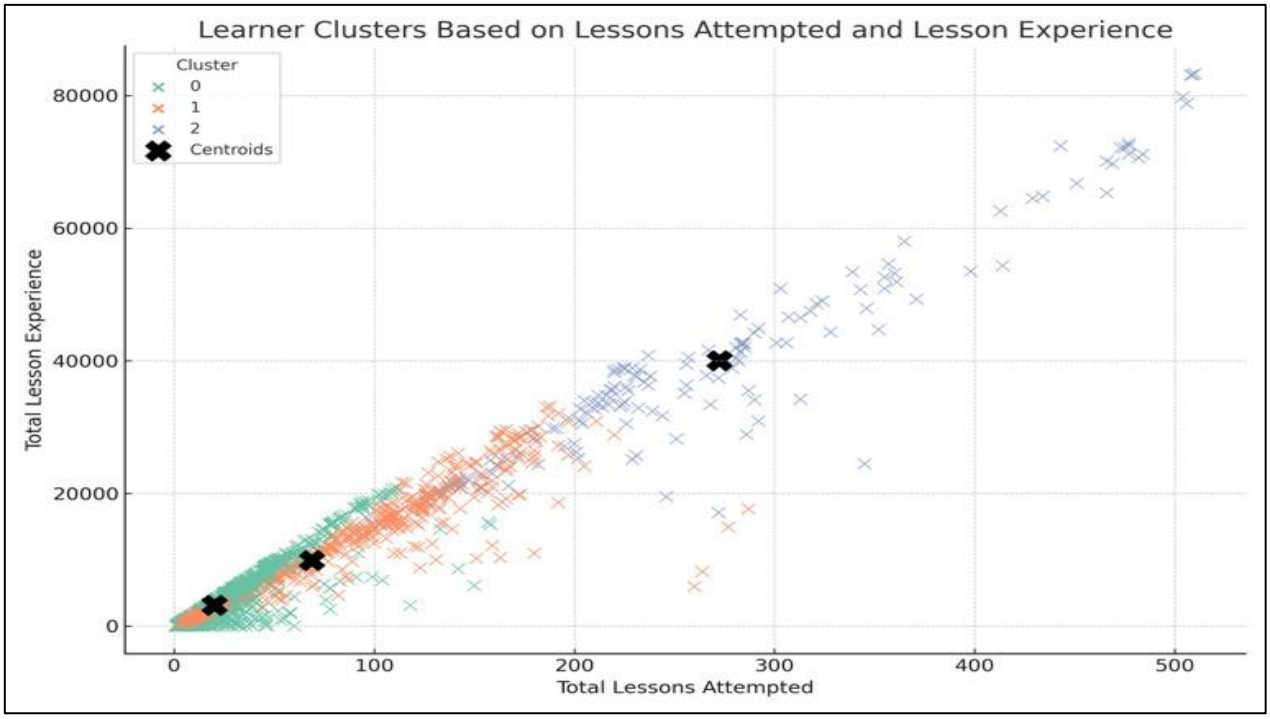

Fig 6 Learner Clusters Based on Total Lessons Attempted and Total Lesson Experience

As illustrated in Figure 6, the scatter plot shows the distribution of learners across the three clusters determined by the K-means procedure. Each point represents a learner and is colored according to the cluster assignment. The cluster centroids are marked with a black X. The plot highlights the differences in learner engagement, thereby aiding the development of personalized teaching methods.

➢ *Cluster Interpretation via Feature Distributions*

To gain a better understanding of the quality and consistency of the identified clusters, we plotted the average values of the most critical learner engagement metrics. These indicators were the number of lesson experiences, submissions, completed lessons, and attempted lessons for the four clusters (see Figure 2). The obtained bar plots display apparent behavioral differences among the groups and indicate how engaged the learners are. For instance, Cluster 3 was characterized by high averages for all metrics.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul298

This identifies it as a highly engaged class, whereas Cluster 0 appears to exhibit the lowest level of activity, likely representing students with low engagement. These results align with the Auto Level scale (0 – 3) and provide a solid foundation for targeted instructional approaches. Moreover, the uniformity of these visual patterns was indicative of the robustness of the clustering model. A good Silhouette Coefficient, along with amicable separation on the PCA plot, supports this.

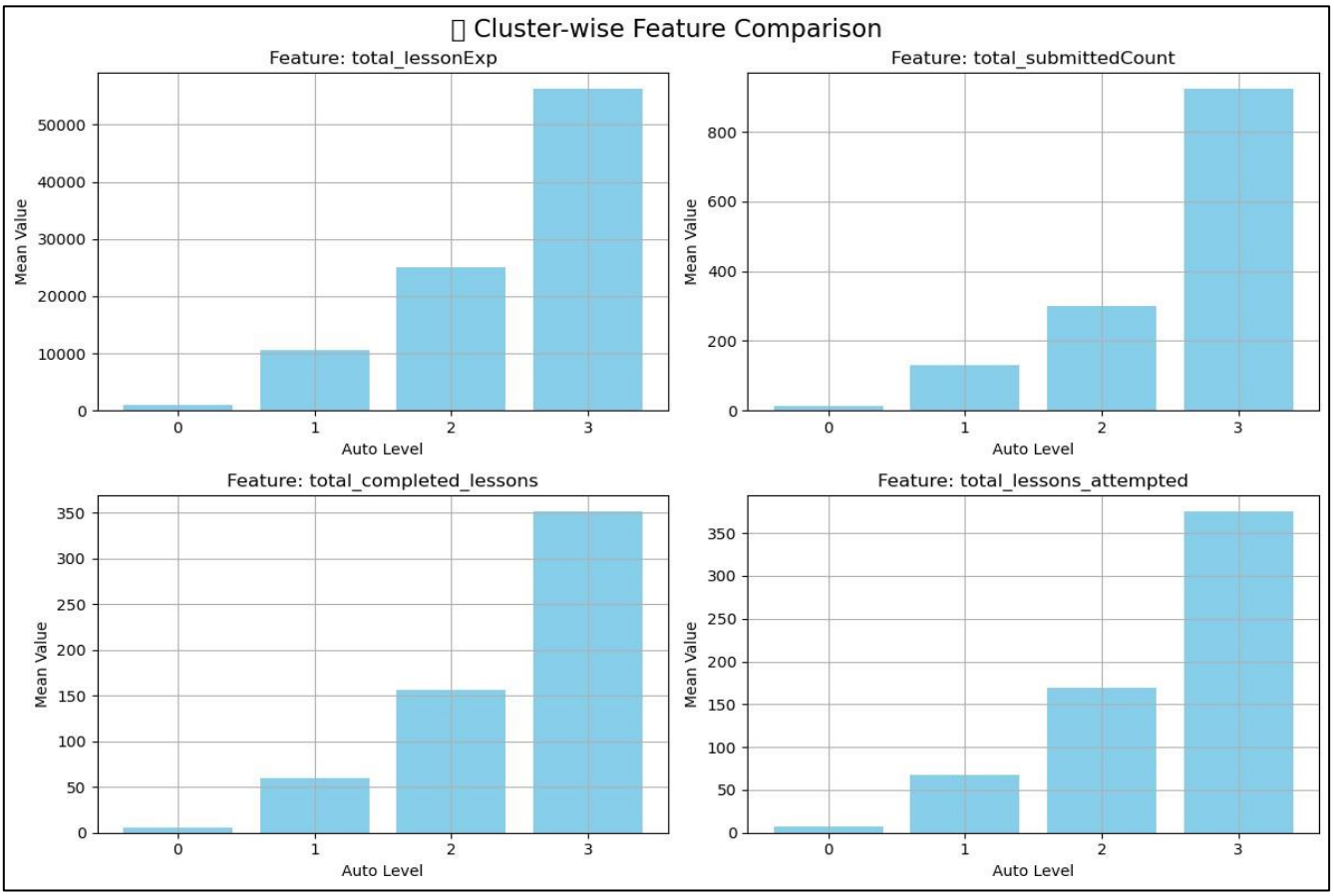

Fig 7 Cluster-Wise Feature Means Comparison (Log Scale)

Figure 7 shows the differences between the four main learning habits of the learner groups. Each bar reflects the average performance of the learners within a cluster. We observed a pattern: learners in Cluster 3 were the most active, completing more lessons, submitting more work, and gaining the most experience. Cluster 2 students showed greater interest, whereas those in Cluster 1 were less active. Cluster 0 contained primarily inactive learners. These patterns help us understand how each group learns differently and form the basis for developing support techniques tailored to their learning needs.

➢ *Statistical Validation of Cluster Differences*

The source dataset comprised several behavioral indicators used to assess engagement in the IT training system. To highlight the features most relevant to clustering methods, we considered four main behavioral variables: total lesson experience (total_lessonExp), total number of assignment submissions (total_submittedCount), total lessons completed (total_completed_lessons), and total lessons attempted (total_lessons_attempted). These covariates were selected based on exploratory data analysis and correlation assessment, which revealed low-to-moderate correlations among the covariates. This implies that all of them can offer distinct information for learner segmentation without significant redundancy.

Table 2 Statistical Validation of Feature Differences Among Clusters via ANOVA

Feature	F-value	p-value
Total Lesson Experience	6475.13	< 0.001
Total Submitted Count	701.79	< 0.001
Total Completed Lessons	7314.10	< 0.001
Total Lessons Attempted	6849.38	< 0.001

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul298

All four variables showed highly significant differences across the four clusters ($p < 0.001$). This confirms that the clusters represent statistically different learner groups based on behavioral variables. To handle missing data, we removed records with null values in any of the four features. This choice preserved the data integrity for clustering. Subsequently, we normalized the features using min-max scaling to adjust their values to a standard range of [0, 1]. This ensures that all features have a uniform influence during the distance-based clustering.

IV. ETHICAL CONSIDERATIONS

This research complied with high ethical standards in analyzing educational data. First, we considered privacy and anonymity and de-identified all student data before analysis to ensure that no member of the research team had access to the actual identities. We used the data in this study with the users' informed consent in accordance with the terms of service between the platform and the learners. Moreover, in the interest of data security, the dataset was locked and stored in a secure location according to the institution's policy. Finally, in terms of fairness and bias, we analyzed the clustering method for potential bias that could induce unfairness or clumping in learner groups. We acknowledge the limitations posed by the dataset size and refrain from drawing universal conclusions.

V. TOOLS AND IMPLEMENTATION ENVIRONMENT

We conducted our study using Python 3.10 on Google Collaboratory (Colab), a cloud-based interactive platform that offers convenient access to computational resources and has many widely used preinstalled libraries. This configuration enables the efficient processing of data-intensive machine learning applications and facilitates collaborative work. We processed and analyzed the data using Pandas and NumPy, and preprocessed and built machine learning models with the help of scikit-learn. Specifically, we used K-means clustering with min-max scaling for feature normalization. Additionally, we utilized Matplotlib and Seaborn for visualization to facilitate the understanding of the results and represent cluster distributions and data trends. Using an open cloud platform ensures that our research is replicable and scalable, adhering to best practices in learning analytics and educational data science. It also encourages transparency and reproducibility, which are central tenets of the academic science community.

VI. RESULTS AND DISCUSSION

➢ *Learner clustering and profiling*

To categorize and analyze learners, we used K-means clustering with k set to 4, as determined by the Elbow Method, on a dataset of 3,384 learners. The clustering process focused on four key behavioral features: the number of finished, submitted, attempted, and completed lessons. Each cluster medoid was assigned an ordinal “Auto Level” ranging from 0 (least engaged) to 3 (most engaged), indicating higher learner activity and estimated skill.

- Auto Level 0: Users exhibit low behavioral interaction and are at risk of disengagement.
- Auto Level 1: These users were described as "low involvement, medium performance" users, with their participation found to be at best intermittent. This may originate from low self-regulation or external motivation.
- Auto Level 2: These students were more involved and exhibited moderate self-regulation, completing many assigned tasks and demonstrating consistent progress. This demonstrates better self-control and dedication.
- Auto Level 3: This category corresponds to highly active learners who presented the highest engagement and performance in all measures, indicating strong self-directedness and academic progression.

The clustering technique is successful in exposing unique learner profiles and could enable a personalized teaching approach based on the engagement levels and learning requirements of students in each group.

Table 3 Learner Profiling Based on Cluster Behavior

Level	Learner Profile Description
0	Low-engagement learners with minimal lesson interaction and task submission
1	Moderately engaged learners with inconsistent performance.
2	Steady participants showing regular engagement and growing autonomy.
3	Highly active and self-regulated learners with superior metrics across all dimensions.

Table 3 shows the learner engagement patterns from K-means clustering analysis. Each level represents a specific pattern of learner behavior, from limited involvement and low task completion at Level 0 to high engagement and advanced self-regulation at Level 3. These profiles clarify the distinctions in learner engagement. They can also provide targeted support to learners with varying levels of engagement and autonomy. The mean comparison plots used a logarithmic scale to examine the behavioral differences among the clusters. This analysis revealed a consistent increase in engagement and performance from clusters 0 to 3. This strong trend across learner levels enhances the validity of the clustering model and supports the rationale for using auto-level categorization.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul298

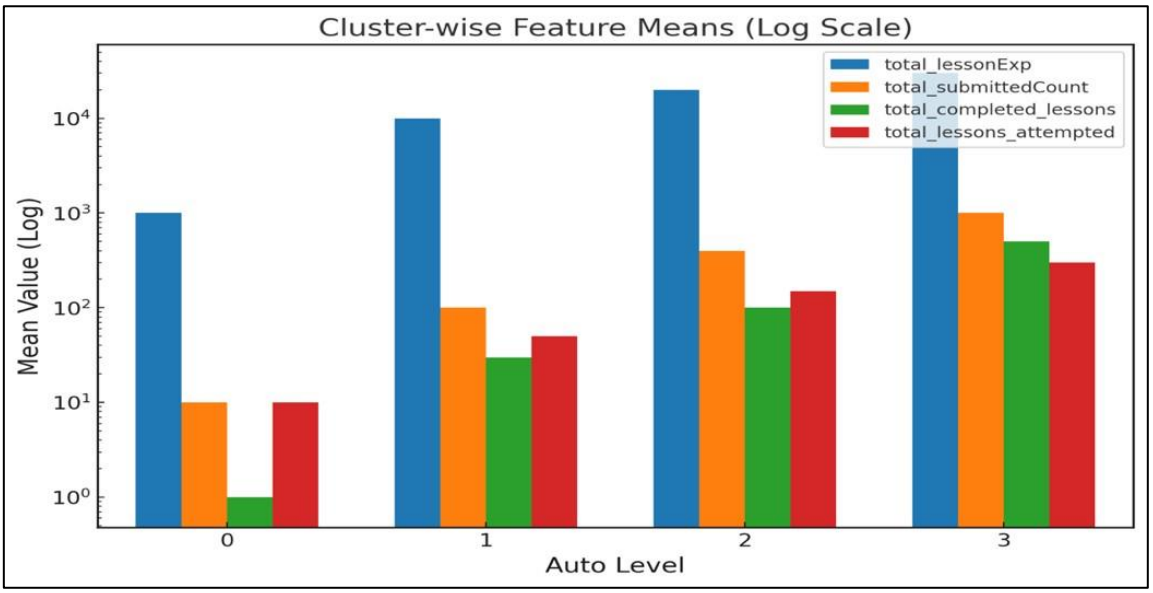

Fig 8 Mean Feature Values Across Clusters

Figure 8 shows the mean of the learned metrics on a logarithmic scale, illustrating the significant variance in the students' dynamics. Driven by the cluster assignments, the mean of each four-test item increases from Level 0 to Level 3, as one would anticipate. The students were categorized into increasingly "involved" profiles. Additional confirmation of the clusters was achieved by investigating the feature distributions. This trend of behavioral characteristics, which were positively correlated with difficulty level from 0 to 3, emerged from our analysis.

➢ *Statistical Validation of Clusters*

The mean differences among clusters regarding salient overall lesson experience features, total submissions, completed lessons, and lessons attempted were identified (one-way ANOVA). The p-values were below 0.001. The statistical findings validate that the clusters are well-defined in terms of competitive segments of learners, demonstrating their unique behaviors. All pair comparisons between clusters, namely, the passive and advanced classes, were significant due to the large differences observed, as shown by the supplementary analyses conducted after the primary analysis. This indicates clear separability among learner profiles. Findings suggest the utility of cluster-based segmentation as one of the techniques for designing targeted instructional designs.

➢ *PCA Visualization of Learner Clusters*

To determine how behavior data were separated based on Clustering, we performed a PCA on the behavior data and projected the high-dimensional data into two main components. The PCA plot representing the similarity relationship among the different groups was presented in Figure 9, further proving the correctness of the K-means cluster analysis. The first two principal components explain 98.90% of the total variance, indicating that almost all the essential student behavior patterns were covered by the four variables identified. This excellent preservation of variance demonstrates the remarkable performance of PCA in capturing class separations of educational data.

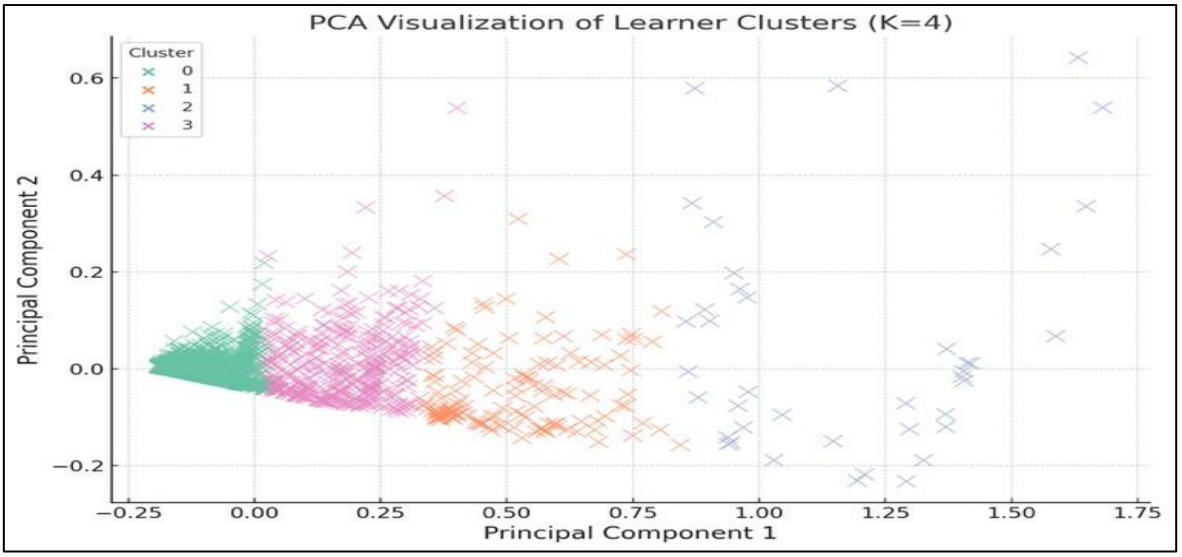

Fig 9 PCA Visualization of Learner Clusters

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul298

All node clusters presented in Figure 9 are associated with the corresponding group along the two-dimensional main component space. This visual split illustrates the model's ability to perform grade-level appropriate learner classification based on engagement and performance characteristics. The separation of different clusters, combined with the high level of explained variance and clustering results, also suggests that the K-means algorithm can capture essential structures in the data. Additionally, the PCA plot reveals the data structure, helping to uncover how variations in student behavior can affect the decisions of educators and researchers about a particular educational strategy, which should be helpful.

- *Adaptive Learning Strategies and Pedagogical Implications*

Classifying learner types through clustering directly impacts educational strategies and methods. For example, learners in the passive group benefit from scaffolded tutorials and gamified onboarding activities, which increase motivation and reduce early disengagement. This aligns with Cognitive Load Theory (Sweller, 1988), emphasizing the importance of structuring information and reducing extraneous load for novice learners. By integrating these insights with the Adaptive Cognitive Enhancement Model (ACEM), the system can automatically identify passive learners and deliver targeted, supportive interventions tailored to their needs.

In contrast, learners in the advanced group thrive when provided with opportunities for autonomous learning and peer mentoring. These methods build on models of self-directed learning**,** social constructivism**,** and communities of practice, which emphasize the importance of collaboration and independence in maintaining engagement. For intermediate learners, adaptive pathways such as guided practice, mastery-based projects, and challenge-based assignments can be used to help them steadily move toward independence.

This personalization demonstrates how clustering techniques can significantly impact retention, motivation, and learning outcomes. It also highlights how teaching strategies can be systematically added to top learning management systems (LMSs). Using ACEM, learner logging data are analyzed in real-time, with profiles and suggested interventions displayed on instructor dashboards. Teachers can then track learner progress and intervene early, creating a continuous cycle of feedback and adjustment. This process enables instructional strategies to be responsive to the changing needs of diverse learner groups, while also remaining scalable and cost-effective.

Table 4 Mapping of Learner Clusters to Pedagogical Strategies and Theoretical Foundations

Learner Cluster	Characteristics	Tailored Instructional Interventions	Theoretical Foundations
Passive	Less engaged, low submissions, high dropout risk	Scaffolded tutorials, step-by-step orientation, and gamified onboarding	Cognitive Load Theory; Motivation Theory
Basic	Inconsistent participation, low-to-moderate performance	Structured feedback, guided practice, mastery-based tasks	Vygotsky's Zone of Proximal Development
Intermediate	Steady participation, moderate autonomy, building competence	Collaborative projects, peer discussions, challenge-based learning	Social Constructivism; Experiential Learning
Advanced	Highly engaged, self-regulated, consistent progress	Self-paced modules, independent projects, and peer mentoring roles	Self-Determination Theory; Communities of Practice

Table 4 presents the mapping of learner clusters to instructional interventions and theoretical foundations. Passive learners benefit from scaffolded support and motivational mechanisms, while advanced learners thrive in autonomous and peer-driven contexts. The Adaptive Cognitive Enhancement Model (ACEM) operationalizes this mapping by processing learner data in real time and presenting actionable profiles to educators through dashboards, thereby enabling responsive and personalized teaching strategies.

VII. LIMITATIONS OF THE STUDY AND FUTURE DIRECTIONS

This paper provides valuable insights into how customizing IT platform education is of great importance, but also, we should also acknowledge some limitations. The basic assumptions of K-means clustering, that every cluster is spherical in shape and that they are sensitive to their initial conditions and outliers, can reduce the accuracy of characterizing learner activities. Also, the missing of cognitive and affective data limits our understanding of the levels of student engagement.

Future studies should consider incorporating both cognitive data (e.g., working memory load) and emotional data (e.g., learners' motivation). Such data is usually collected through survey questionnaires, physiological measurements, or sentiment analysis of discussion forum posts. Integrating this data with behavioral information will be very helpful in developing more detailed learner profiles and adaptive strategies. As the research was carried out within a single IT training environment, its findings may not be easily replicable to other systems, domains, or learner groups. Future studies should explore possible alternative or hybrid clustering approaches to overcome these challenges and to determine the level of effectiveness of the proposed model in different learning environments in the long run, considering its decay effects. Such attempts will enhance the

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul298

effectiveness and usefulness of adaptive technologies in digital learning.

VIII. CONCLUSION

The results of this paper reveal that K-means clustering is a realistic approach to categorizing learners into different groups on an IT-based learning platform and, therefore, facilitates scalable and personalized learning experiences. The paper focuses on various pedagogical strategies through ACEM that turn learner groups into opportune interventions. This aspect will allow educators to design personalized learning experiences, thus enhancing student engagement. The study shows that unsupervised machine learning has the potential to improve personalized education through the utilization of data-driven methods that are more progressive than traditional standardized methods.

Although the research is limited by an unbalanced dataset and relies mainly on the behavior features of a single platform, etc., it lays a solid base for scalable adaptive learning models. The mentioned limitations will be addressed in future studies by incorporating cognitive and emotional data, experimenting on other platforms, and developing hybrid clustering models. The findings of the research may be employed to support personalized digital learning environments with adaptive data-driven practices. The results provide practical insights for teachers, as well as the platforms' designers, to promote adaptive and learner-centered learning environments.

- *Declaration of Conflicting Interest*

The authors declared that there is no potential conflict of interest regarding the research, authorship, or publication of this article.

- *Declaration of Generative AI in the Writing Process*

We also acknowledge the support of ChatGPT in refining the language of our early drafts. Its use was limited to improving clarity and grammar, and all material was carefully reviewed and revised by the authors before submission.

REFERENCES

[1]. ABDULNASSAR, A. A., & NAIR, L. R. (2023). Performance analysis of K-means with modified initial centroid selection algorithms and the developed Kmeans9+ model. *Measurement: Sensors, 25*, 100666. https://doi.org/10.1016/j.measen.2023.100666

[2]. AGHA, D., MEGHJI, A. F., & BHATTI, S. (2023). Clusters of success: Unpacking academic trends with K-means clustering in education. *VFAST Transactions on Software Engineering, 11*(4), 15–31. https://doi.org/10.21015/vtse.v11i4.1633

[3]. ALAWI, S. J. S., SHAHARANEE, I. N. M., & JAMIL, J. M. (2023). Clustering student performance data using k-means algorithms. *Journal of Computational Innovation and Analytics, 2*(1), 41–55. https://doi.org/10.32890/jcia2023.2.1.3

[4]. AMALIA BAHARUDDIN, H. M. Z. (2022). Mining e-learning interactions using K-means clustering. *AIP Conference Proceedings, 2644*(1), 030013. https://doi.org/10.1063/5.0104447

[5]. BREMPONG, O. W., JR. (2021). K-means clustering algorithm for students' selection and performance prediction. *International Journal of Scientific & Technology Research, 10*(10), 35–40. https://www.ijstr.org/finalprint/oct2021/K-means-Clustering-Algorithm-For-Students-Selection-And-Performance-Prediction.pdf

[6]. CHANG, W., JI, X., LIU, Y., XIAO, Y., CHEN, B., LIU, H., & ZHOU, S. (2020). Analysis of university students' behavior based on a fusion K-means clustering algorithm. *Applied Sciences, 10*(18), 6566. https://doi.org/10.3390/app10186566

[7]. CHATTERJEE, S., & CHOUDHURY, S. J. (2025). Exploring structural components in autoencoder-based data clustering. *Engineering Applications of Artificial Intelligence, 140*, 109562. https://doi.org/10.1016/j.engappai.2024.109562

[8]. DE MARSICO, M., STERBINI, A., & TEMPERINI, M. (2013). A framework to support social-collaborative personalized e-learning. In *Human-Computer Interaction. Applications and Services* (pp. 351–360). Springer. https://doi.org/10.1007/978-3-642-39262-7_40

[9]. ERKAN, E., SILIK, S., & CANSIZ, S. (2024). Uncovering engagement profiles of young learners in K–8 education through learning analytics. *Journal of Learning Analytics, 11*(1), 101–115. https://doi.org/10.18608/jla.2024.8133

[10]. 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 Second International Conference on Knowledge Discovery and Data Mining (pp. 226–231). https://dl.acm.org/doi/10.5555/3001460.3001507

[11]. FADIEIEVA, L. O. (2023). Adaptive learning: A cluster-based literature review (2011–2022). *Educational Technology Quarterly, 3*, 319–366. https://doi.org/10.55056/etq.613

[12]. GARCIA, M., LOPEZ, R., & HERNANDEZ, F. (2019). Linking clustering results to teaching strategies: A framework for educational data mining. *Computers & Education, 140*, 103604. https://doi.org/10.1016/j.compedu.2019.103604

[13]. HENDERI, H., SUNARYA, A., ZAKARIA, Z., NURMIKA, S. L., & ASMAINAH, N. (2020). Evaluation model of students' learning outcomes using the k-means algorithm. *Journal of Physics: Conference Series, 1477*, 022027. https://doi.org/10.1088/1742-6596/1477/2/022027

[14]. IFENTHALER, D., & YAU, J. Y.-K. (2020). Utilizing learning analytics to support study success in higher education: A systematic review. https://doi.org/10.1007/s11423-020-09788-z

[15]. 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

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul298

[16]. KEM, DR. (2022). Personalized and adaptive learning: Emerging learning platforms in the era of digital and smart learning. *International Journal of Social Science and Human Research, 5*(2). https://doi.org/10.47191/ijsshr/v5-i2-02

[17]. KURNIAWAN, R. W. D., & HERMAWAN, A. (2023). *Personalized student learning mechanisms using K-means and Kmedoids clustering algorithms based on individual preferences*. *International Journal of Computer Applications, 185*(46), 13–19. https://doi.org/10.5120/ijca2023923274

[18]. LIU, H., ZHENG, M., & WU, Y. (2023). Enhancing computer programming education with large language models: A study on effective prompt engineering. *Journal of Educational Computing Research*. https://arxiv.org/abs/2407.05437

[19]. LUO, Y. (2024). Revolutionizing education with AI: The adaptive cognitive enhancement model (ACEM) for personalized cognitive development. *Applied and Computational Engineering, 82*, 71–76. https://doi.org/10.54254/2755-2721/82/20240929

[20]. MOHD TALIB, N. I., ABD MAJID, N. A., & SAHRAN, S. (2023). Identification of student behavioral patterns in higher education using K-means clustering and support vector machine. *Applied Sciences, 13*(5), 3267. https://doi.org/10.3390/app13053267

[21]. NGUYEN, T., & DO, P. (2021). Clustering-based approaches for learner modeling in online education. Computers & Education, 168, 104211. https://doi.org/10.1016/j.compedu.2021.104211

[22]. OYELADE, O. J., OLADIPUPO, O. O., & OBAGBUWA, I. C. (2010). Applying the k-means clustering algorithm for predicting students' academic performance. *International Journal of Computer Science and Information Security, 7*(1), 292–295. https://www.researchgate.net/publication/45900764_

[23]. OZDEMIR, D., OPSETH, H. M., & TAYLOR, H. (2020). Leveraging learning analytics for student reflection and course evaluation. *Journal of Applied Research in Higher Education, 12*(1), 27–37. https://doi.org/10.1108/JARHE-11-20180253

[24]. PAPAMITSIOU, Z., & ECONOMIDES, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. Educational Technology & Society, 17(4), 49–64. https://www.researchgate.net/publication/267510046

[25]. RIZVI, S., RIENTIES, B., ROGATEN, J., & KIZILCEC, R. F. (2022). Beyond one-size-fits-all in MOOCs: Variation in learning design and learners' persistence in different cultural and socioeconomic contexts. *Computers in Human Behavior, 126*, 106973. https://doi.org/10.1016/j.chb.2021.106973

[26]. ROUSSEEUW, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

[27]. SHIRIN MOJARAD, M., ESSA, S., MOJARAD, S., BAKER, R. S., & MCGRAW, B. (2018). Data-driven learner profiling based on clustering student behaviors: Learning consistency, pace, and effort. https://learninganalytics.upenn.edu/ryanbaker/ITS_2018_Shirin_final.pdf

[28]. SIRISHA, N., MAGESWARI, P., MOHAN RAJ, V., KUMAR, S., PRIYA, R. V., & ANANTHI, S. (2025). Emotion-centric AI-driven engagement systems for adaptive learning environments: Personalized knowledge acquisition and cognitive skill enhancement. In Proceedings of the International Conference on Sustainability Innovation in Computing and Engineering (ICSICE 2024) (pp. 528–540). Atlantis Press. https://doi.org/10.2991/978-94-6463-718-2_46

[29]. SWELLER, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

[30]. TALIB, N. I. M., MAJID, N. A. A., & SAHRAN, S. (2023). Identifying student behavioral patterns in higher education using K-means clustering and support vector machines. *Applied Sciences, 13*(5), 3267. https://doi.org/10.3390/app13053267

[31]. VANKAYALAPATI, R., GHUTUGADE, K., VANNAPURAM, R., & PRASANNA, B. (2021). K-means algorithm for Clustering of learners' performance levels using machine learning techniques. *Revue d'Intelligence Artificielle, 35*, 99–104. https://doi.org/10.18280/ria.350112

[32]. VYGOTSKY, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press. https://home.fau.edu/musgrove/web/vygotsky1978.pdf

[33]. WEI, D., & ZHANG, Z. (2023). K-means clustering algorithm based on improved density peak. In *Proceedings of the 2023 3rd International Conference on Bioinformatics and Intelligent Computing (BIC'23)* (pp. 105–109). Association for Computing Machinery. https://doi.org/10.1145/3592686.3592706

[34]. XU, D., & TIAN, Y. (2015). A comprehensive survey of clustering algorithms. Annals of Data Science, 2(2), 165–193. https://doi.org/10.1007/s40745-015-0040-1