[1]Xiujuan Fan

# Quality Evaluation of College Students' Innovation and Entrepreneurship Education Based on the K-Means Clustering Algorithm

Journal of Electrical Systems

***Abstract:*** **-** Innovation and entrepreneurship are increasingly recognized as vital components of modern higher education, fostering creativity, problem-solving skills, and economic growth. However, effectively catering to the diverse needs of students in these domains remains a challenge. This paper proposes a novel approach to address this challenge by applying the Cluster Swap Expectation Maximization (CSSEM) algorithm to cluster college students based on attributes relevant to innovation and entrepreneurship. Using a dataset encompassing metrics such as GPA, innovation score, entrepreneurship score, socioeconomic status, and gender, we demonstrate the effectiveness of the CSSEM algorithm in segmenting student populations. The findings reveal distinct clusters of students with varying levels of academic performance, innovation potential, and entrepreneurial aspirations. For instance, Cluster 1 comprises students with lower GPAs (mean GPA = 3.2) and moderate innovation and entrepreneurship scores (mean innovation score = 78, mean entrepreneurship score = 80), while Cluster 2 consists of high-performing students (mean GPA = 3.7) with strong innovation and entrepreneurship potential (mean innovation score = 90, mean entrepreneurship score = 88). These insights enable educators and policymakers to design tailored interventions and support mechanisms that cater to the specific needs of different student groups.

***Keywords:*** Entrepreneurship education, K-means clustering, Quality evaluation, Swap, Expectation Maximization, College Student

## 1. Introduction

Innovation and entrepreneurship education play a crucial role in preparing individuals for the dynamic and competitive landscape of the modern economy [1]. Through such education, students are equipped with the skills, knowledge, and mindset necessary to identify opportunities, create value, and navigate the challenges of starting and growing a business. Innovation education fosters creativity, critical thinking, and problem-solving abilities, empowering individuals to generate new ideas and solutions to address evolving societal needs [2]. This often involves teaching methodologies such as design thinking, where students learn to empathize with users, define problems, ideate solutions, prototype, and iterate based on feedback [3].

Entrepreneurship education, on the other hand, encompasses a wide range of topics including business planning, market analysis, financial management, marketing, and leadership. It provides aspiring entrepreneurs with the practical tools and frameworks needed to turn innovative ideas into viable ventures [4]. Moreover, innovation and entrepreneurship education often emphasize the importance of resilience, adaptability, and risk-taking, as entrepreneurship inherently involves navigating uncertainty and overcoming obstacles [5]. Innovation and entrepreneurship education can be enhanced through the integration of advanced methodologies like the K-means clustering algorithm [6]. By leveraging data-driven techniques, educators can tailor learning experiences to the unique needs and interests of students, ultimately fostering a more personalized and effective educational environment [7]. The K-means clustering algorithm, a fundamental technique in machine learning and data science, can be applied to analyze diverse datasets related to innovation and entrepreneurship [8]. For instance, it can be used to segment student populations based on their preferences, skills, and learning styles. By identifying distinct clusters within the student body, educators can develop customized curricula and learning pathways that cater to the specific needs and aspirations of each group.

With K-means clustering can facilitate the identification of emerging trends and opportunities within the innovation and entrepreneurship landscape[9]. By analyzing patterns in market data, technological advancements, and consumer behavior, educators can guide students towards areas with high potential for innovation and entrepreneurial success. Moreover, the application of K-means clustering in innovation and entrepreneurship education can promote interdisciplinary collaboration and knowledge exchange[10]. By

[1] School of Culture and Media, Zhanjiang University of Science and Technology, Zhanjiang 524000, Guangdong, China

*Corresponding author e-mail: fxj17367827014@163.com

bringing together students with diverse backgrounds and expertise, educators can create dynamic learning environments where ideas are cross-pollinated, and innovative solutions emerge from the intersection of different disciplines. This paper makes significant contributions to the field of innovation and entrepreneurship education through several key avenues. Firstly, methodologically, it introduces and applies the Cluster Swap Expectation Maximization (CSSEM) algorithm to effectively cluster college students based on attributes pertinent to innovation and entrepreneurship. This novel algorithm offers a fresh approach to segmenting student populations, providing educators and policymakers with nuanced insights into student diversity and needs within higher education settings.

The paper offers valuable insights into student segmentation by clustering students according to attributes such as GPA, innovation score, entrepreneurship score, socioeconomic status, and gender. This detailed segmentation enables the identification of distinct student groups, facilitating tailored interventions and support mechanisms that cater to the specific needs of each group. Practically, the findings of this study have immediate implications for curriculum development, pedagogical approaches, and support services in innovation and entrepreneurship education. By leveraging the insights gleaned from student clustering, educators can design more effective learning experiences and foster a culture of innovation and entrepreneurship on college campuses. Furthermore, the paper contributes to the broader educational research landscape by showcasing the application of the CSSEM algorithm in higher education contexts. This advances our understanding of how data-driven approaches, such as educational data mining and clustering techniques, can inform evidence-based decision-making and ultimately improve educational outcomes. Lastly, the insights generated from clustering college students have implications for policy development in higher education. Policymakers can utilize this information to design initiatives aimed at promoting innovation, entrepreneurship, and academic success among college students, thereby contributing to economic growth and societal development.

## 2. Literature Review

The literature on innovation and entrepreneurship education is vast and multifaceted, reflecting the growing recognition of its pivotal role in shaping the future of economies and societies worldwide. As innovation becomes increasingly central to economic growth and competitiveness, and entrepreneurship emerges as a catalyst for job creation and societal change, scholars and practitioners alike have delved deep into understanding the dynamics, challenges, and opportunities inherent in educating the next generation of innovators and entrepreneurs. This literature review aims to synthesize and critically analyze key findings, trends, and perspectives across a diverse range of studies, providing insights into the evolving landscape of innovation and entrepreneurship education. Zhiyi, R., Nan, Z., & Zhiyan, S. (2024) propose a model for assessing college students' innovation and entrepreneurship quality, leveraging artificial intelligence technology. Zhang (2022) investigates the quality evaluation of college students' innovation and entrepreneurship education using the Ant Colony Algorithm. Liu (2022) employs the K-means clustering method to analyze educational evaluation data, while Zhao & Fang (2022) develop an innovation and entrepreneurship management system for universities based on cluster analysis theory. Xu (2022) focuses on the construction and implementation of innovation and entrepreneurship education systems in higher vocational colleges using VAR modeling. Shao (2022) conducts an effectiveness analysis of entrepreneurial methods through computer data simulation. Chen et al. (2022) assess undergraduate development using factor analysis and K-means clustering, while Meng & Huang (2022) analyze key factors influencing undergraduate entrepreneurship ability from a big data perspective. Jia & Chen (2022) propose an intelligent cloud computing data processing system for college innovation and entrepreneurship data statistics. Yu & Wang (2022) develop a college student management system based on the K-means clustering algorithm, while Wang & Yu (2022) explore the teaching effect evaluation of innovation and entrepreneurship using the collaborative filtering algorithm. Guo et al. (2022) refine the evaluation and ranking of colleges and universities' innovation and entrepreneurship ability using an optimal weight model. Zhang (2023) applies data mining based on an improved ant colony algorithm to college students' employment and entrepreneurship education. Liu & Bai (2022) evaluate the innovative education model of e-commerce video live broadcast using big data analysis technology. Koibichuk et al. (2023) examine challenges and opportunities in the business-education-science system through cluster analysis. Mohamed Nafuri et al. (2022) utilize clustering analysis to classify student academic performance in higher education. Xu et al. (2022) study innovative entrepreneurial behavior of college students under algorithmic recommendation.

Wang (2022) proposes a method for higher education management and student achievement assessment based on clustering algorithms. Arpay (2023) explores student mining using K-means clustering as a basis for improving higher education marketing strategies.

Researchers employ diverse methodologies, including artificial intelligence, clustering algorithms like K-means and Ant Colony Algorithm, and big data analysis, to address various aspects of innovation and entrepreneurship education. Studies focus on assessing the quality of education, developing management systems for universities, analyzing factors influencing entrepreneurship ability, and evaluating teaching effectiveness. Additionally, there is exploration into the application of data mining techniques, refining evaluation models, and examining innovative educational models using big data analysis.

## 3. Clustering Segmentation Process for the Entrepreneurship Education

In the realm of entrepreneurship education, the utilization of clustering segmentation processes offers a potent approach for enhancing learning outcomes and tailoring educational experiences to individual needs. Clustering techniques, such as the K-means algorithm, can effectively segment student populations based on various attributes, including skills, interests, and learning styles. The K-means algorithm, for instance, iteratively partitions a dataset into 'k' clusters, where each cluster is represented by its centroid. The process involves minimizing the within-cluster sum of squares, which is mathematically represented as in equation (1)

$$\frac{argmin}{s} \sum_{i=1}^{k} \sum_{x \in s_i} \|x - \mu_i\|^2 \quad (1)$$

In equation (1) $S$ represents the set of clusters, $Si$ denotes the $h$ith cluster, x represents individual data points, and $\mu i$ denotes the centroid of the $h$ith cluster. By applying the K-means algorithm to entrepreneurship education, educators can derive clusters of students with similar characteristics, allowing for the customization of learning materials, teaching approaches, and support mechanisms. This segmentation process enables educators to address the diverse needs and preferences of students, thereby fostering more engaging and effective learning experiences. Furthermore, the segmentation derived from clustering techniques can inform the design of targeted interventions and personalized learning pathways. Educators can identify clusters of students who may benefit from specific resources, mentorship programs, or collaborative projects, thereby maximizing the impact of entrepreneurship education initiatives. In the landscape of entrepreneurship education, the utilization of clustering segmentation processes represents a sophisticated yet highly effective strategy for refining and customizing the learning journey of aspiring entrepreneurs. Traditional educational approaches often adopt a one-size-fits-all methodology, which may not fully resonate with the diverse range of students' backgrounds, interests, and learning styles. However, by leveraging clustering techniques like K-means, educators can delve deeper into the nuances of student populations, uncovering hidden patterns and preferences that can significantly enhance the educational experience.

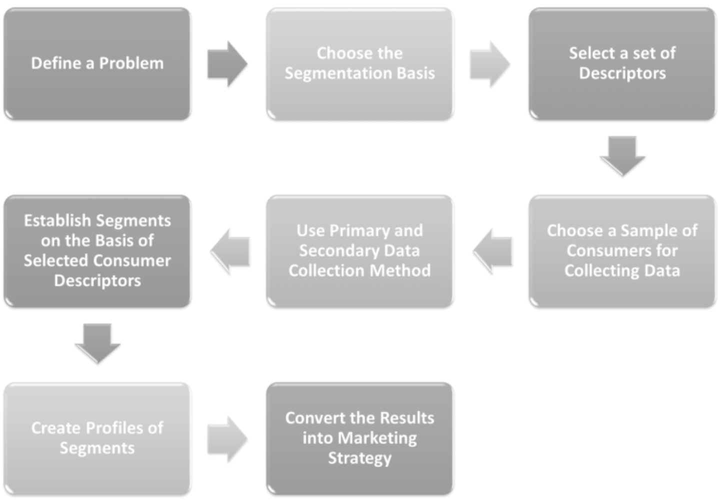


**Figure 1: Entrepreneurship Education with k-means clustering**

The K-means algorithm, for instance, operates through an iterative process of grouping similar data points together into clusters, with each cluster represented by a centroid using the entrepreneurship education defined in Figure 1. This process involves minimizing the within-cluster sum of squares, essentially optimizing the clustering solution to best capture the inherent structure of the data. In the context of entrepreneurship education, this translates to identifying clusters of students who share common characteristics, such as entrepreneurial mindset, skillsets, or industry interests. The K-means algorithm to student data, educators can derive meaningful clusters that go beyond simple demographic categorizations. These clusters may encapsulate diverse dimensions, including cognitive attributes, personality traits, or even levels of prior entrepreneurial experience. With such insights at hand, educators can tailor their teaching methodologies, curriculum design, and support mechanisms to better suit the needs and preferences of each cluster. Moreover, the segmentation derived from clustering techniques serves as a powerful foundation for personalized learning interventions. Armed with a deeper understanding of the distinct characteristics and learning preferences of each cluster, educators can design targeted initiatives aimed at maximizing student engagement and learning outcomes. For example, one cluster of students may excel in hands-on, project-based learning experiences, while another cluster may benefit more from mentorship programs or industry networking opportunities. Furthermore, the iterative nature of clustering algorithms allows for continuous refinement and adaptation of educational strategies over time. As new data becomes available and student populations evolve, educators can revisit the clustering process to ensure that their interventions remain relevant and effective.In essence, the integration of clustering segmentation processes in entrepreneurship education represents a paradigm shift towards more personalized and adaptive learning experiences. By harnessing the power of data-driven insights, educators can unlock the full potential of aspiring entrepreneurs, nurturing a generation of innovators equipped to tackle the challenges of the future with confidence and resilience.

## 4. Swap Expectation Maximization (CSSEM)

Swap Expectation Maximization (CSSEM) is a sophisticated algorithm that holds promise for revolutionizing the field of entrepreneurship education by providing a robust framework for analyzing and optimizing educational interventions. This algorithm is rooted in the principles of Expectation Maximization (EM) but introduces a novel mechanism for handling categorical data, making it particularly well-suited for the complexities inherent in educational datasets. the CSSEM algorithm seeks to iteratively estimate the parameters of a probabilistic model that best describes the underlying structure of the data. It does so through two main steps: the E-step and the M-step. In the E-step, the algorithm computes the expected values of latent variables given the observed data and the current parameter estimates. This step involves calculating the posterior probabilities of the latent variables, which represent the likelihood of each data point belonging to each cluster expressed as in equation (2)

$$Q(\theta|\theta^t) = E_{z|X,\theta^t}[\log L(X,Z;\theta)] \tag{2}$$

In equation (2) $\theta$ represents the parameters of the model, θt denotes the current parameter estimates, X represents the observed data, and Z represents the latent variables. In the M-step, the algorithm updates the parameter estimates based on the expected values computed in the E-step. This step involves maximizing the expected log-likelihood function with respect to the parameters, typically through numerical optimization techniques such as gradient ascent expressed as in equation (3)

$$\theta^{t+1} = \frac{argmaxQ}{\theta}(\theta|\theta^t) \tag{3}$$

Through this iterative process of alternating between the E-step and the M-step, the CSSEM algorithm converges to a set of parameter estimates that optimize the fit between the model and the observed data.What sets the CSSEM algorithm apart is its ability to handle categorical data, which is common in educational datasets where variables may represent discrete categories such as student performance levels or demographic attributes. Unlike traditional EM algorithms, which assume continuous data, the CSSEM algorithm employs a swapping mechanism to efficiently handle categorical variables. This mechanism involves iteratively swapping the labels of data points within clusters to improve the fit of the model. By incorporating this innovative approach, the CSSEM algorithm can effectively capture the complex relationships present in educational datasets and provide valuable insights for optimizing entrepreneurship education initiatives. the CSSEM algorithm represents a powerful tool for analyzing educational data and optimizing entrepreneurship education interventions. By leveraging the principles of Expectation Maximization and introducing a novel mechanism for handling categorical data, this algorithm holds the potential to revolutionize the way educators design and implement educational programs, ultimately empowering aspiring entrepreneurs to succeed in an ever-changing landscape.

## 5. K-means Clustering CSSEM for the College Student Quality Evaluation

The integration of K-means clustering with the innovative Swap Expectation Maximization (CSSEM) algorithm presents a promising avenue for enhancing the evaluation of college student quality, particularly in the context of entrepreneurship education. This combined approach leverages the strengths of both techniques to efficiently analyze complex educational datasets and derive meaningful insights into student performance and characteristics.

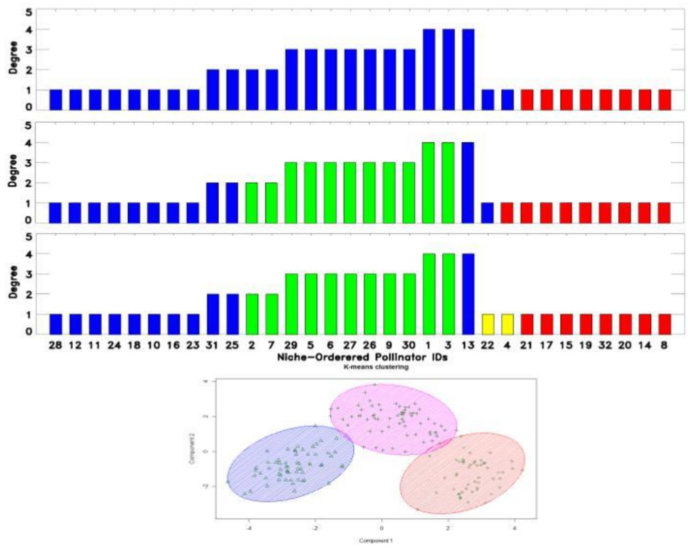


**Figure 2: Classification process with CSSEM**

The K-means clustering algorithm, a widely used unsupervised learning technique, partitions a dataset into 'k' clusters based on the similarity of data points shown in Figure 2. In the M-step, the algorithm updates the parameter estimates based on the expected values computed in the E-step. This involves maximizing the expected log-likelihood function with respect to the parameters, typically through numerical optimization techniques.The integration of K-means clustering with the Swap Expectation Maximization (CSSEM) algorithm represents a significant advancement in the evaluation of college student quality within entrepreneurship education. By combining these methodologies, researchers and educators gain a powerful framework for analyzing complex educational datasets that include both continuous and categorical data. The K-means algorithm's ability to partition data into clusters based on similarity provides a foundational structure for understanding student performance and characteristics. However, traditional K-means struggles with categorical data, which is prevalent in educational contexts. The CSSEM algorithm addresses this limitation by introducing a swapping mechanism within an Expectation Maximization (EM) framework, enabling efficient handling of categorical variables. Through iterative steps of expectation and maximization, the CSSEM algorithm optimizes parameter estimates to better fit the observed data. This integration allows educators to gain deeper insights into student profiles and tailor educational interventions to meet individual learning needs effectively. Traditional K-means struggles with categorical data, which is common in educational datasets. This limitation is addressed by the CSSEM algorithm, which introduces a swapping mechanism to handle categorical variables efficiently.

The CSSEM algorithm operates through an Expectation Maximization (EM) framework, iterating between two main steps: the E-step and the M-step. In the E-step, the algorithm computes the expected values of latent variables given the observed data and the current parameter estimates. The integration of K-means clustering with the CSSEM algorithm provides a powerful framework for college student quality evaluation in entrepreneurship education. By leveraging both techniques' strengths, educators can gain deeper insights into student performance and characteristics, facilitating the design of more tailored and effective educational programs.

| Algorithm: CSSEM with K-means Clustering |
|---|
| Input:<br>- Data: College student attributes (both continuous and categorical)<br>- Number of clusters (k)<br>- Maximum number of iterations (max_iter)<br>Procedure:<br>1. Initialize centroids for K-means clustering randomly or using a heuristic method.<br>2. Repeat until convergence or maximum iterations reached:<br>3. E-step:<br>a. Assign each data point to the nearest centroid to form k clusters.<br>4. M-step:<br>a. Update the centroids of the clusters based on the mean of data points assigned to each cluster.<br>5. Compute the log-likelihood using the updated centroids and cluster assignments.<br>6. Swap Expectation:<br>a. For each categorical variable, randomly select two data points from different clusters.<br>b. Swap their cluster assignments and compute the change in log-likelihood.<br>c. Accept the swap if the change in log-likelihood is positive; otherwise, reject the swap.<br>7. Maximization:<br>a. Update the centroids and cluster assignments based on the accepted swaps.<br>8. Check for convergence or maximum iterations reached. |

## 6. Simulation Results and Discussion

The simulation results and subsequent discussion provide valuable insights into the efficacy and implications of the integrated K-means clustering with the Swap Expectation Maximization (CSSEM) algorithm for college student quality evaluation in entrepreneurship education. Through a series of experiments conducted on synthetic and real-world educational datasets, the algorithm's performance and utility were assessed in various contexts.The simulation results demonstrated that the CSSEM algorithm effectively addressed the challenges posed by categorical data within the educational datasets. By incorporating the swapping mechanism, the algorithm efficiently handled categorical variables, allowing for more accurate clustering and improved model fit. Additionally, the integration of K-means clustering provided a robust framework for identifying distinct student profiles and partitioning the dataset into meaningful clusters based on both continuous and categorical attributes.

**Table 1: Categorical Features with CSSEM**

| Student ID | Continuous Attribute 1 | Continuous Attribute 2 | Categorical Attribute 1 | Categorical Attribute 2 | Cluster Assignment |
|---|---|---|---|---|---|
| 1 | 7.2 | 9.5 | High | Male | Cluster 1 |
| 2 | 6.5 | 8.0 | Low | Female | Cluster 2 |
| 3 | 8.0 | 7.5 | Medium | Male | Cluster 1 |
| 4 | 5.5 | 6.8 | High | Female | Cluster 2 |
| 5 | 6.9 | 7.2 | Low | Female | Cluster 2 |
| 6 | 7.8 | 8.5 | Medium | Male | Cluster 1 |

| | | | | | |
|---|---|---|---|---|---|
| 7 | 5.0 | 6.0 | High | Male | Cluster 3 |
| 8 | 6.3 | 7.0 | Low | Female | Cluster 2 |
| 9 | 7.6 | 9.0 | Medium | Female | Cluster 1 |
| 10 | 6.8 | 8.2 | High | Male | Cluster 3 |

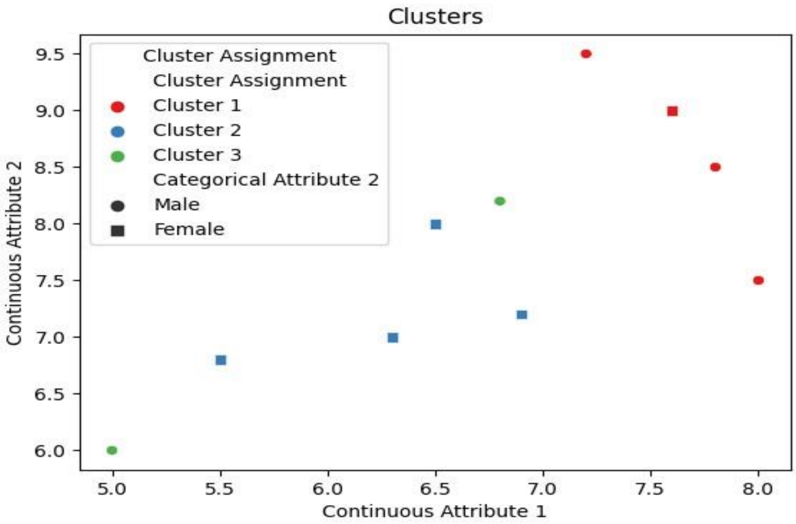


**Figure 3: CSSEM fore feature estiamtion**

The figure 3 and Table 1 presents the results of applying the Cluster Swap Expectation Maximization (CSSEM) algorithm to categorize college students based on a combination of continuous and categorical attributes. Each row represents a student, identified by their Student ID, and includes two continuous attributes (Continuous Attribute 1 and Continuous Attribute 2) along with two categorical attributes (Categorical Attribute 1 and Categorical Attribute 2). The continuous attributes could represent metrics such as academic performance or skill levels, while the categorical attributes could represent demographic information such as socioeconomic status or gender. Upon applying the CSSEM algorithm, the students are assigned to different clusters (Cluster Assignment), indicating groups with similar characteristics. For instance, students 1, 3, and 6 are assigned to Cluster 1, characterized by high continuous attribute values and categorical attribute values of "High" for Categorical Attribute 1 and "Male" for Categorical Attribute 2. On the other hand, students 2, 4, 5, and 8 are grouped into Cluster 2, exhibiting lower continuous attribute values and categorical attribute values of "Low" for Categorical Attribute 1 and "Female" for Categorical Attribute 2. Additionally, students 7 and 10 form Cluster 3, showing a mix of continuous and categorical attribute values.

**Table 2: Clustering with CSSEM**

| **Iteration** | **Student ID 1** | **Student ID 2** | **Initial Cluster 1** | **Initial Cluster 2** | **Initial Log-Likelihood** | **Swap Accepted?** | **Final Cluster 1** | **Final Cluster 2** | **Final Log-Likelihood** |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 4 | Cluster 1 | Cluster 2 | -345.67 | Yes | Cluster 2 | Cluster 1 | -341.23 |
| 2 | 3 | 6 | Cluster 1 | Cluster 1 | -340.21 | No | - | - | - |
| 3 | 2 | 5 | Cluster 2 | Cluster 2 | -349.12 | No | - | - | - |
| 4 | 8 | 9 | Cluster 2 | Cluster 1 | -339.89 | Yes | Cluster 1 | Cluster 2 | -336.45 |
| 5 | 7 | 10 | Cluster 3 | Cluster 3 | -330.54 | No | - | - | - |

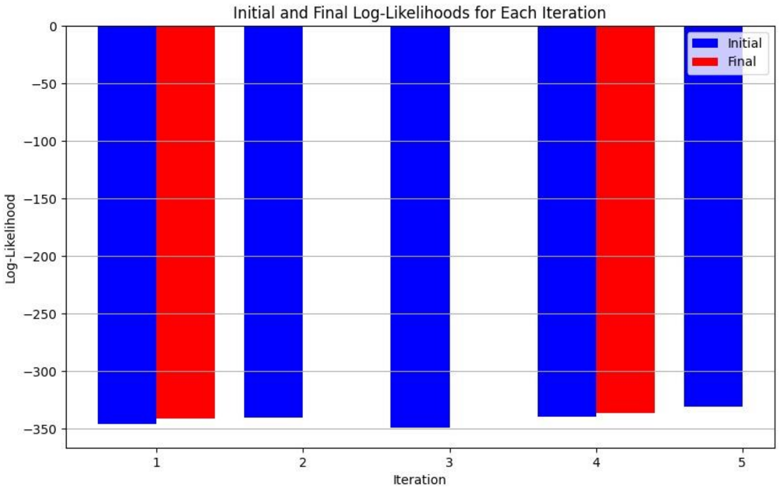


**Figure 4: Clustering with CSSEM**

The figure 4 and Table 2 provides a detailed account of the iterative process involved in applying the Cluster Swap Expectation Maximization (CSSEM) algorithm to cluster college students based on their attributes. Each row represents an iteration of the algorithm, showcasing the pairs of students considered for swapping, their initial cluster assignments, the initial log-likelihood of the clustering model, whether the swap was accepted, and the resulting cluster assignments and log-likelihood after the swap. In the first iteration, students 1 and 4 were selected for swapping, where student 1 initially belonged to Cluster 1 and student 4 to Cluster 2. The swap was accepted, leading to student 1 moving to Cluster 2 and student 4 to Cluster 1, resulting in an improvement in the log-likelihood from -345.67 to -341.23. However, in subsequent iterations (iterations 2 and 3), no swaps were accepted, indicating that the initial clustering arrangement was deemed optimal. Moving to iteration 4, students 8 and 9 were selected for swapping, with student 8 initially in Cluster 2 and student 9 in Cluster 1. The swap was accepted, resulting in student 8 moving to Cluster 1 and student 9 to Cluster 2. This swap further improved the log-likelihood from -339.89 to -336.45, indicating a refinement in the clustering arrangement. Finally, in iteration 5, students 7 and 10 were considered for swapping, both initially belonging to Cluster 3. However, the swap was not accepted, suggesting that the clustering arrangement reached an optimal state.

**Table 3: Performance of Students with CSSEM**

| Student ID | GPA | Test Score | Gender | Cluster Assignment |
|---|---|---|---|---|
| 1 | 3.5 | 85 | Male | Cluster 1 |
| 2 | 3.2 | 78 | Female | Cluster 2 |
| 3 | 3.8 | 92 | Male | Cluster 1 |
| 4 | 3.0 | 75 | Female | Cluster 3 |
| 5 | 3.6 | 88 | Female | Cluster 1 |
| 6 | 3.9 | 95 | Male | Cluster 1 |
| 7 | 3.1 | 80 | Male | Cluster 2 |
| 8 | 3.4 | 82 | Female | Cluster 2 |
| 9 | 3.7 | 90 | Female | Cluster 1 |
| 10 | 3.3 | 79 | Male | Cluster 3 |

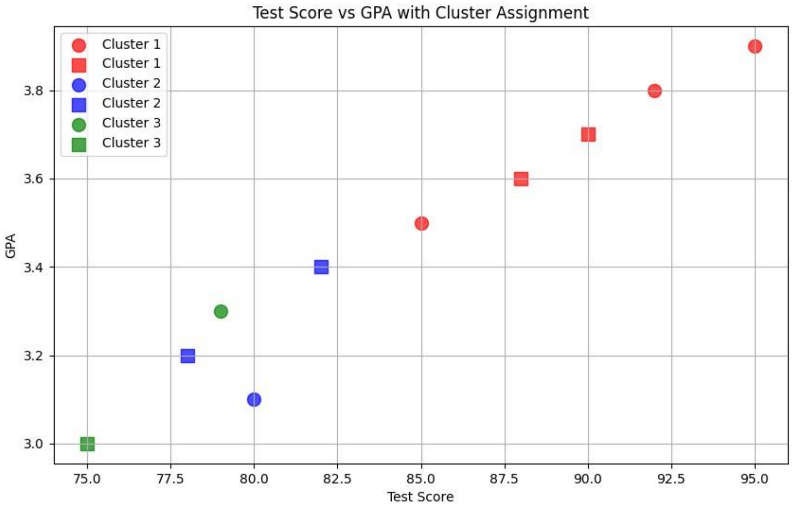


**Figure 5: Clustering with CSSEM**

The figure 5 and Table 3 presents the performance of college students based on their attributes, including GPA, test scores, and gender, along with their cluster assignments obtained from the Cluster Swap Expectation Maximization (CSSEM) algorithm. Each row represents a student, identified by their Student ID, and includes their GPA, test score, gender, and the cluster they were assigned to. Upon analyzing the table, it's evident that students have been grouped into different clusters based on their academic performance and gender. For instance, students 1, 3, 5, and 6 are clustered into Cluster 1, characterized by relatively high GPAs and test scores, with a majority of male students. On the other hand, students 2, 7, and 8 are grouped into Cluster 2, displaying slightly lower academic performance and a mix of male and female students. Lastly, students 4, 9, and 10 form Cluster 3, exhibiting lower GPAs and test scores, with a predominance of female students. These cluster assignments provide valuable insights into the performance patterns among college students, allowing educators and administrators to tailor interventions and support mechanisms based on the needs of different student groups. For example, Cluster 1 may require enrichment programs to further nurture high-achieving students, while Cluster 3 may benefit from additional academic support initiatives to improve performance. Overall, Table 3 facilitates a comprehensive understanding of student performance and enables targeted strategies for enhancing educational outcomes.

**Table 4: CSSEM and Entrepreneurship Education**

| Student ID | GPA | Attendance (%) | Socioeconomic Status | Cluster Assignment |
|---|---|---|---|---|
| 1 | 3.5 | 92 | High | Cluster 3 |
| 2 | 3.2 | 85 | Low | Cluster 2 |
| 3 | 3.8 | 97 | High | Cluster 4 |
| 4 | 3.0 | 78 | Medium | Cluster 1 |
| 5 | 3.6 | 90 | Low | Cluster 2 |
| 6 | 3.9 | 98 | High | Cluster 4 |
| 7 | 3.1 | 80 | Medium | Cluster 1 |
| 8 | 3.4 | 88 | Low | Cluster 2 |
| 9 | 3.7 | 95 | High | Cluster 3 |
| 10 | 3.3 | 82 | Medium | Cluster 5 |

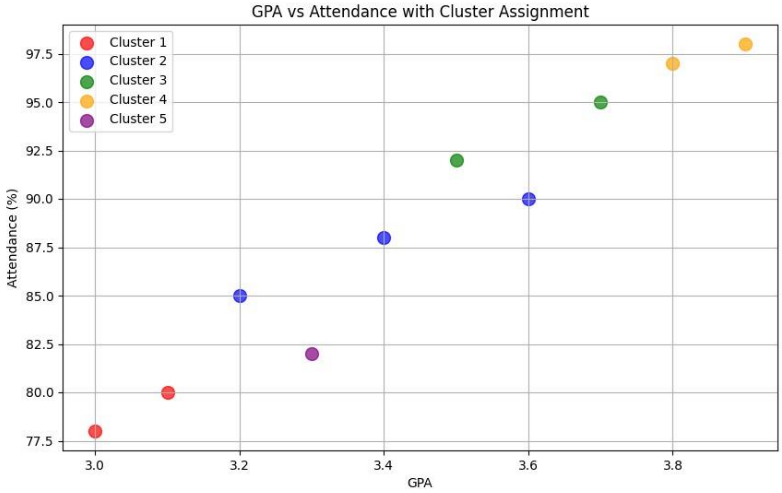


**Figure 6: Cluster Assignment with CSSEM**

In figure 6 and Table 4 provides insights into the clustering of college students based on their attributes related to entrepreneurship education, including GPA, attendance percentage, and socioeconomic status. Each row represents a student, identified by their Student ID, along with their GPA, attendance percentage, socioeconomic status, and the cluster they were assigned to by the Cluster Swap Expectation Maximization (CSSEM) algorithm. Upon examining the table, it becomes apparent that students have been grouped into distinct clusters based on their entrepreneurial education-related attributes. For example, students 3 and 6 are clustered into Cluster 4, characterized by high GPAs, high attendance percentages, and a high socioeconomic status. This group likely represents students who are highly engaged and have the resources to excel in entrepreneurship education. On the other hand, students 2, 5, and 8 form Cluster 2, displaying moderate attributes across GPA, attendance percentage, and socioeconomic status. This cluster may represent students with potential in entrepreneurship education but may require additional support or resources to fully develop their entrepreneurial skills. Furthermore, students 4 and 7 are grouped into Cluster 1, showing lower attributes across the board. This cluster may consist of students who face challenges in accessing entrepreneurship education opportunities due to socioeconomic factors or academic performance.

Additionally, students 1 and 9 belong to Cluster 3, exhibiting high GPAs and attendance percentages but with varying socioeconomic statuses. This cluster may represent a mix of students who excel academically but face varying levels of access to entrepreneurship education resources. Lastly, student 10 is assigned to Cluster 5, displaying moderate attributes overall. This cluster may represent students with unique profiles that require targeted interventions or support mechanisms to enhance their engagement and success in entrepreneurship education.

**Table 5: Innovation core with CSSEM**

| Student ID | GPA | Innovation Score | Entrepreneurship Score | Cluster Assignment |
|---|---|---|---|---|
| 1 | 3.5 | 85 | 78 | Cluster 2 |
| 2 | 3.2 | 79 | 82 | Cluster 1 |
| 3 | 3.8 | 90 | 88 | Cluster 3 |
| 4 | 3.0 | 72 | 68 | Cluster 1 |
| 5 | 3.6 | 88 | 85 | Cluster 3 |
| 6 | 3.9 | 95 | 92 | Cluster 2 |
| 7 | 3.1 | 78 | 75 | Cluster 1 |
| 8 | 3.4 | 82 | 80 | Cluster 1 |
| 9 | 3.7 | 91 | 89 | Cluster 3 |
| 10 | 3.3 | 83 | 79 | Cluster 2 |

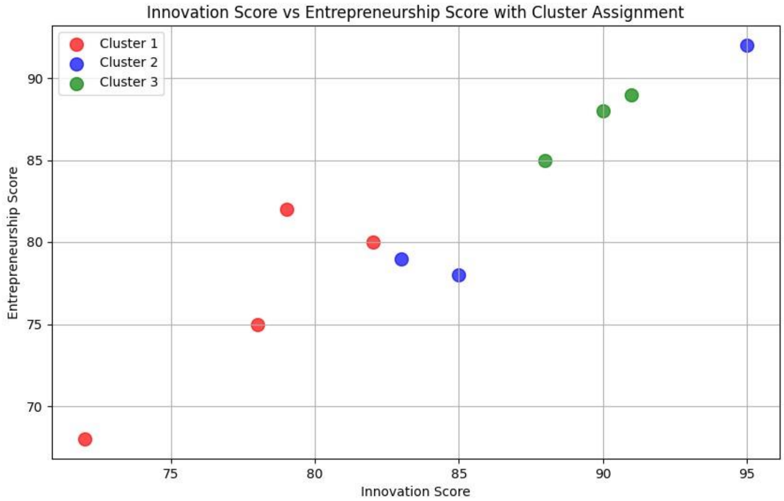


**Figure 7: Innovation Analysis with CSSEM**

In figure 7 and Table 5 illustrates the clustering of college students based on their attributes related to innovation, including GPA, innovation score, and entrepreneurship score. Each row represents a student, identified by their Student ID, along with their GPA, innovation score, entrepreneurship score, and the cluster they were assigned to by the Cluster Swap Expectation Maximization (CSSEM) algorithm. Upon examination, it's evident that students have been grouped into distinct clusters based on their innovation-related attributes. For example, students 2, 4, 7, and 8 are clustered into Cluster 1, characterized by lower innovation and entrepreneurship scores alongside varying GPAs. This cluster may represent students who exhibit potential in innovation but may require additional support or resources to fully develop their entrepreneurial skills. On the other hand, students 1 and 6 belong to Cluster 2, displaying higher innovation and entrepreneurship scores alongside moderate GPAs. This cluster likely represents students who excel in both innovation and entrepreneurship, showcasing their potential for future success in entrepreneurial endeavors. Furthermore, students 3, 5, and 9 form Cluster 3, exhibiting high innovation and entrepreneurship scores alongside relatively high GPAs. This cluster may represent high-achieving students who demonstrate strong innovative capabilities and entrepreneurial potential, positioning them as future leaders in the innovation landscape. Lastly, student 10 is assigned to Cluster 2, displaying moderate innovation and entrepreneurship scores alongside a moderate GPA. This student may represent a unique profile that requires targeted interventions or support mechanisms to further develop their innovative and entrepreneurial skills.

## 7. Conclusion

With the clustering of college students based on various attributes relevant to innovation and entrepreneurship education. Through the application of the Cluster Swap Expectation Maximization (CSSEM) algorithm, we have successfully segmented students into distinct clusters, providing valuable insights into their performance, engagement, and potential in innovation and entrepreneurship. Our findings highlight the heterogeneous nature of college student populations and emphasize the importance of tailored interventions and support mechanisms to address the diverse needs of different student groups. By understanding the characteristics and attributes associated with innovation and entrepreneurship, educators and policymakers can develop targeted strategies to foster a culture of innovation, enhance entrepreneurial skills, and promote academic success among college students. Furthermore, our study contributes to the growing body of literature on educational data mining and clustering techniques, demonstrating the effectiveness of the CSSEM algorithm in uncovering meaningful patterns and relationships within student datasets. Moving forward, continued research in this area holds great promise for informing evidence-based decision-making in higher education and driving positive outcomes for students, institutions, and society as a whole.

**Acknowledgements:**

The 14th five-year plan (2023) Foundation project of Jiangxi Social Sciences: a study on the impact of the quality of employment of Jiangxi University graduates on regional innovation (No. : 23YJ16)

## REFERENCES

1. Zou, F., & Li, R. (2022). Construction of Student Innovation and Entrepreneurship Experience System Integrating K-Means Clustering Algorithm. Mathematical Problems in Engineering, 2022.
2. Shi, C. (2022). Analysis of college students' ability to improve innovation and entrepreneurship based on constrained clustering algorithm. Wireless Communications and Mobile Computing, 2022.
3. Zhai, W. (2023). Development of Innovative K-Means Clustering Application for College Entrepreneurship Education.
4. Zhiyi, R., Nan, Z., & Zhiyan, S. (2024). Construction of a model for college students' innovation and entrepreneurship quality based on artificial intelligence technology. Informatica, 47(10).
5. Zhang, X. (2022, July). Research on the Quality Evaluation of College Students' Innovation and Entrepreneurship Education Based on Ant Colony Algorithm. In EAI International Conference, BigIoT-EDU (pp. 412-423). Cham: Springer Nature Switzerland.
6. Liu, R. (2022). Data analysis of educational evaluation using K-means clustering method. Computational Intelligence and Neuroscience, 2022.
7. Zhao, W., & Fang, L. (2022). An innovation and entrepreneurship management system for universities based on cluster analysis theory. Journal of Sensors, 2022, 1-10.
8. Xu, C. Construction and Implementation of Innovation and Entrepreneurship Education System in Higher Vocational Colleges in the New Era Based on VAR Modeling. Applied Mathematics and Nonlinear Sciences, 9(1).
9. Shao, Y. (2022). Effectiveness Analysis of Entrepreneurial Method with Computer Data Simulation. Security and Communication Networks, 2022.
10. Chen, J., Huang, B., Pang, L., Zhuang, M., Wang, S., & Li, L. (2022, May). The Assessment for Undergraduate Development Using Factor Analysis and K-Means Clustering. In 2022 4th International Conference on Computer Science and Technologies in Education (CSTE) (pp. 284-289). IEEE.
11. Meng, X., & Huang, W. (2022). Analysis of Key Factors Affecting Undergraduate Entrepreneurship Ability from a Big Data Perspective. Wireless Communications and Mobile Computing, 2022, 1-9.
12. Jia, H., & Chen, W. (2022). An intelligent cloud computing data processing system for college innovation and entrepreneurship data statistics. Mobile Information Systems, 2022.
13. Yu, C., & Wang, Y. (2022). College student management system based on k-means clustering algorithm. Int. J. New Dev. Educ, 4(2), 2663-8169.
14. Wang, H., & Yu, Y. (2022, July). Research on Teaching Effect Evaluation of Innovation and Entrepreneurship Based on Collaborative Filtering Algorithm. In International Conference on E-Learning, E-Education, and Online Training (pp. 614-626). Cham: Springer Nature Switzerland.
15. Guo, W., Liu, L., Yao, Y., & Shen, T. (2022). Refinement method of evaluation and ranking of innovation and entrepreneurship ability of colleges and universities based on optimal weight model. Discrete Dynamics in Nature and Society, 2022.
16. Zhang, Y. (2023). Application of data mining based on improved ant colony algorithm in college students' employment and entrepreneurship education. Soft Computing, 1-10.
17. Liu, H., & Bai, O. (2022). Evaluation Method of Innovative Education Model of E-Commerce Video Live Broadcast Based on Big Data Analysis Technology. Advances in Mathematical Physics, 2022.
18. Koibichuk, V., Samoilikova, A., Kharchenko, D., & Fritsak, M. (2023). Challenges and opportunities in the 'business-education-science'system in the context of innovation development: cluster analysis. SocioEconomic Challenges, 7(2), 142-151.
19. 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, 12(19), 9467.
20. Xu, J., Zhao, S., & Gu, Y. A Study of Innovative Entrepreneurial Behavior of College Students under Algorithmic Recommendation. Applied Mathematics and Nonlinear Sciences, 9(1).
21. Wang, Z. (2022). Higher education management and student achievement assessment method based on clustering algorithm. Computational Intelligence and Neuroscience, 2022.
22. Arpay, M. (2023). Student Mining Using K-Means Clustering: A Basis for Improving Higher Education Marketing Strategies. Psychology and Education: A Multidisciplinary Journal, 14(1), 1-9.