Zuzana Kubincová · Dalila Durães ·
María Cruz Sánchez-Gómez · Paulo Novais ·
Loreto Lancia · Maria Angela Pellegrino
Editors

Methodologies and Intelligent Systems for Technology Enhanced Learning, Workshops - 14th International Conference

Springer

Editors
Zuzana Kubincová
Comenius University in Bratislava
Bratislava, Slovakia

Dalila Durães
Universidade do Minho
Braga, Portugal

María Cruz Sánchez-Gómez
Didactics, Organization and Research Method
University of Salamanca
Salamanca, Salamanca, Spain

Paulo Novais
Departamento de Informática
University of Minho
Braga, Portugal

Loreto Lancia
Department of Life, Health and Health
and Environmental Sciences
University of L'Aquila
Coppito, Italy

Maria Angela Pellegrino
ISISLabSstecca F, Secondo Piano
Università degli Studi di Salerno
Fisciano, Salerno, Italy

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-031-85560-3 ISBN 978-3-031-85561-0 (eBook)
https://doi.org/10.1007/978-3-031-85561-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Education, science, research and technology deployment in all areas of life are considered to be the basic pillars of the knowledge society. The International Conference in Methodologies and Intelligent Systems for Technology Enhanced Learning (mis4TEL) serves as a forum for experts from all these fields, including not only education, information technology or computer science, but also such disciplines as psychology, medicine, social sciences, etc. It encourages multidisciplinary research and discussion on technology-enhanced learning promoting new intelligent and creative solutions for formal as well as informal learning and all types of learners. In addition to technological solutions, the technology-enhanced learning approach can be fostered by novel methods coming from different fields of research and from diverse communities also including "fragile users," like children, elderly people or people with special needs.

The annual appointment of mis4TEL established itself as a consolidated fertile forum where scholars and professionals from the international community, with a broad range of expertise in the TEL field, share results and compare experiences. The conference program also features four selected workshops which aim to provide participants with the opportunity to present and discuss novel research ideas on emerging topics complementing the main conference.

This volume presents the papers that were accepted for the following workshops of mis4TEL 2024: *Artificial Intelligence for Education (Ai4Ed), Technology Enhanced Learning in Nursing Education* (NURSING), *Technology Enhanced Learning for Future Citizens* (TEL4FC), and *Integration of emerging technologies into education and training* (ETELT).

All papers underwent a peer-review selection: each paper was assessed by two/three different reviewers, from an international panel of each workshop. A total of 18 quality papers, with authors coming from various countries, have been selected for the workshops and included in the present volume.

This edition of the conference is organized by the BISITE Research Group of the University of Salamanca (Spain). We would like to thank all the contributing authors, the members of the program committee, the reviewers, the sponsors and the organizing committee for their hard and highly valuable work. Thanks for your help–mis4TEL 2024 would not exist without your contribution.

Zuzana Kubincová
Dalila Durães
María Cruz Sánchez-Gómez
Paulo Novais
Loreto Lancia
Maria Angela Pellegrino

Organization of mis4TEL 2024

http://www.mis4tel-conference.net/

General Chair

Christothea Herodotou	The Open University, UK

Technical Program Chairs

Sofia Papavlasopoulou	Norwegian University of Science and Technology, Norway
Carlos Santos	University of Aveiro, Portugal

Workshop Chairs

Zuzana Kubincová	Comenius University in Bratislava, Slovakia
Dalila Durães	University of Minho, Portugal

Scientific Chairs

Marcelo Milrad	Linnaeus University, Sweden
Nuno Otero	University of Greenwich, UK

Publicity Chair

María Cruz Sánchez-Gómez	University of Salamanca, Spain

Steering Committee Representatives

Pierpaolo Vittorini	University of L'aquila, Italy
Rosella Gennari	Free University of Bozen-Bolzano, Italy
Tania Di Mascio	University of L'aquila, Italy

Marco Temperini	Sapienza University of Rome, Italy
Fernando De la Prieta	University of Salamanca, Spain

Organizing Committee

Fernando De la Prieta (Chair)	University of Salamanca, Spain
Sara Rodríguez González	University of Salamanca, Spain
Javier Prieto Tejedor	University of Salamanca, Spain
Ricardo S. Alonso Rincón	AIR Institute, Spain
Alfonso González Briones	University of Salamanca, Spain
Pablo Chamoso Santos	University of Salamanca, Spain
Javier Parra	University of Salamanca, Spain
Belén Pérez Lancho	University of Salamanca, Spain
Ana Belén Gil González	University of Salamanca, Spain
Ana De Luis Reboredo	University of Salamanca, Spain
Angélica González Arrieta	University of Salamanca, Spain
Angel Luis Sánchez Lázaro	University of Salamanca, Spain
Emilio S. Corchado Rodríguez	University of Salamanca, Spain
Guillermo Hernández	University of Salamanca, Spain
Liliana Durón	University of Salamanca, Spain
Laura Grande	University of Salamanca, Spain
Carmen Losada	AIR Institute, Spain
Albano Carrera González	AIR Institute, Spain
Marta Plaza Hernández	University of Salamanca, Spain
Beatriz Bellido	University of Salamanca, Spain
María Alonso	University of Salamanca, Spain
Diego Valdeolmillos	AIR Institute, Spain
Carlos Álvarez	AIR Institute, Spain

Categorizing Students of the MathE Platform: a Fuzzy Clustering Perspective*

Gabriel Leite[1,2][0009−0008−3273−0450], Beatriz Flamia Azevedo[1,2][0000−0002−8527−7409], Maria F. Pacheco[1,2][0000−0001−7915−0391], Florbela P. Fernandes[1,2][0000−0001−9542−4460], and Ana I. Pereira[1,2][0000−0003−3803−2043]

Research Centre in Digitalization and Intelligent Robotics (CeDRI), Laboratório Associado para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC), Instituto Politécnico de Bragança, 5300-252 Bragança, Portugal
{gabriel.leite, beatrizflamia, pacheco, fflor, apereira}@ipb.pt

Abstract. Active learning and technology integration offer enhanced student engagement and adaptive learning, accommodating diverse preferences. This work uses fuzzy clustering method to analyze the data of students who answer questions on the MathE platform. To do this, the Fuzzy c-means algorithm was used, which allows flexibility and adaptability in the clustering partitioning, especially in situations where data elements may exhibit overlapping characteristics or belong to multiple categories. Thereby, two datasets are considered: the first is composed of 121 students who answered questions from the Vector Space subtopic, and the second dataset comprises the answers of 297 students who answered to any topic or subtopic of the platform. The results show that the fuzzy clustering method is appropriate for analyzing the student's data since most students are highly associated with more than one cluster. Besides, the findings can support the formulation of intervention strategies to improve the student's academic achievement.

Keywords: e-learning · active learning · higher education.

1 Introduction

Developing learning capabilities and the learning process heavily relies on students' active involvement in their education [2]. Research using case methodologies suggests active learning strategies to increase lecture attendance and engagement and foster expert attitudes within the discipline, enhancing students' capabilities [2, 10, 12].

* This work has been supported by FCT Fundação para a Ciência e Tecnologia within the R&D Units Project Scope UIDB/00319/2020, UIDB/05757/2020 (DOI: 10.54499/UIDB/057 57/2020), UIDP/05757/2020 (DOI: 10.54499/UIDP/05757/2020) and Erasmus Plus KA2 within the project 2021-1-PT01-KA220-HED-000023288. Beatriz Flamia Azevedo is supported by FCT Grant Reference SFRH/BD/07427/2021.

Implementing active learning encounters obstacles, and digital tools is one strategy to enhance it. E-learning tools have surged in popularity due to their ability to offer students and teachers a flexible and engaging approach to studying and learning. The inclusion of student's learning preferences may influence the development of critical thinking skills and their engagement in their learning process. One of the crucial objectives of Artificial Intelligence in education is to provide personalized learning guidance to students based on their individual learning process, preferences, or characteristics [14].

This context led to the MathE platform (mathe.pixel-online.org), an online educational system launched in 2019. Its purpose is to help students overcome challenges associated with college-level Mathematics or those aiming for a more profound understanding of various mathematical subjects, all at their pace [4, 7]. Inside the iMath project, the MathE Platform is being improved to become even more interactive and gain intelligence for decision-making (more information in imath.pixel-online.org). Thus, it is expected that in the near future, the platform can guide the student's learning process autonomously, instead of in a randomized manner, as it currently is. To do this, it is necessary to recognize patterns in the data obtained so far. Thus, this work evaluates the student's behavior when answering questions.

Effectively categorizing student performance is challenging, necessitating meticulous attention to diverse factors like subjectivity, disparities, and engagement, among others. Furthermore, a single student may fit into multiple categories simultaneously in many instances. In such scenarios, fuzzy clustering emerges as a valuable tool, particularly adept at accommodating instances where each element can possess memberships in multiple clusters to varying degrees. This adaptability renders fuzzy clustering particularly advantageous when data elements demonstrate ambiguity or uncertainty concerning cluster assignments.

In the literature it is possible to find several works exploring fuzzy clustering methods to analyze students' behavior. The work [15] explores possible academic courses with significant contributions to academic performance and predicts students' graduating class of degree using a combination between Fuzzy c-means, k-means, and the Adaptive Neuro Fuzzy Inference System to classify the students based on academic performance. The results can support the monitoring groups of students with similar performance levels and defining intervention strategies to enhance academic achievement. In turn, [16] explored the application of fuzzy clustering algorithms for assessing psychological health among undergraduate students. The outcomes present a versatile and accurate approach to understanding the mental health scenario, paving the way for innovative avenues in tailored support and interventions. Similarly [13] investigated the psychological factors affecting students' academic performance. Therefore, the Heuristic Fuzzy c-means Clustering Algorithm is applied to analyze college students' stress levels, psychological well-being, and academic performance detection.

This paper uses fuzzy clustering methods to categorize the MathE students based on their performance when answering the MathE platform questions. Specifically, the Fuzzy c-means algorithm is employed to analyze data collected through the platform [5], considering the number of questions each student answers and the type of answer (correct or incorrect). The results of this study offer valuable insights into evaluating student behavior on the platform and supporting research on optimizing resource utilization for student support.

This paper is organized as follows: after the introduction, Sect. 2 describes the Fuzzy c-means algorithms and the methodology utilized to define the optimum number of clustering partitioning. Afterwards, Sect. 3 presents the datasets utilized in this work. Section 4 presents and discusses the obtained results. Finally, the main conclusion and future direction are presented in Sect. 5.

2 Fuzzy Clustering Method

Clustering can be defined as grouping elements with similarities in the same group and those with dissimilarities in other different groups based on certain criteria [1]. Two common types of clustering are *crisp clustering* and *fuzzy clustering*. Crisp clustering assigns each data point to exactly one cluster, and there are clear boundaries between clusters frontiers. In crisp clustering, each element is associated with a cluster centroid that is closest to in terms of some distance metric. The k-means algorithm is one of the simplest and most efficient clustering algorithms proposed in the literature [6].

On the other hand, performing hard assignments of elements to clusters is not feasible in complex datasets, mainly in uncertainty or ambiguity in data [1]. A fuzzy clustering algorithm can extract such overlapping structures. Fuzzy clustering allows for the possibility that a data element belongs to multiple clusters simultaneously with varying degrees of membership. The Fuzzy c-means (FCM) is a well-known fuzzy clustering algorithm [1, 6]. It assigns membership values to data elements, indicating the degree to which they belong to each cluster. Unlike crisp clustering, in fuzzy clustering, there are no strict boundaries between clusters; instead, there's a gradual transition in membership values, so each data element contributes to each cluster based on its similarity to the centroid [1].

The FCM formulation minimizes the SSE function that captures the similarity between data elements and cluster centroids, as represented in Equation (1),

$$SEE = \sum_{k=1}^{K} \sum_{x_i \in C_k} \mu_{x_{ik}}^{m} ||x_i - c_k||^2, \tag{1}$$

where x_i is a element of the cluster C_k, c_k is the centroid of C_k, m is the fuzzy fuzziness exponent, used to control the degree of fuzzy overlap, with $m > 1$. And, $\mu_{x_{ik}}$ is the degree of membership of the x_ith element in the kth cluster, K is the number of clusters. Fuzzy overlap refers to the fuzzy boundaries between clusters, the number of data elements with significant membership in more than one cluster. The membership values, $\mu_{x_{ik}}$, satisfy the following constraints: each data element's membership values sum up to 1, and each data element's membership value in each cluster lies between 0 and 1.

The FCM algorithm works similarly to k-means where the algorithm minimizes the SSE iteratively, followed by updating $\mu_{x_{ik}}$ and c_k. In each iteration, the membership values are updated using the equations defined in Equation (2).

$$\mu_{x_{ik}} = \frac{1}{\sum_{j=1}^{K} \left(\frac{x_i - c_k}{x_i - c_j} \right)^{\frac{2}{m-1}}}, \text{ and } c_k = \frac{\sum_{x_i \in C_k} \mu_{x_{ik}}^m x_i}{\sum_{x_i \in C_k} \mu_{x_{ik}}}. \tag{2}$$

This process continues until the convergence of centroids occurs [1]. As in k-means, the FCM algorithm required the previous indication of the optimum number of clusters. For this, the Elbow method [9] is considered, which uses the within-cluster sum of square to defines the optimum number of clusters.

3 Datasets

The MathE platform answers dataset comprises 833 questions, organized into 14 topics and 24 subtopics, provided by 372 higher education students from 8 countries (Portugal, Lithuania, Italy, Ireland, Romania, Russia, Spain, and Slovenia), as described in [5]. New questions can be added to the platform at any time, so some questions have a much higher number of answers than others.

Although in the first version of the platform, the questions were divided into 2 difficult levels (Basic and Advanced), that is not an effective way to categorize the questions, as already was demonstrated in several works [4, 7]. So, based on the results of [4], in this work, the questions were reorganized in three levels of difficulty considering the methodology presented in [4, 7]. Two datasets were considered; first, the answers were collected by the Vector Space subtopic since it is one of the most accessed subtopics of the MathE platform. After that, the complete dataset was considered for a global analysis of the students' behavior.

The dataset of Vector Space is composed of 2749 answers distributed among 40 questions provided by 146 students of different nationalities. Whereas the whole data involves 372 students that provide 9470 answers distributed among 833 questions, as detailed in [3, 5]. Table 1 presents more details about the datasets considered.

Table 1: Questions answered according to the type of answer and difficulty level

Question Difficulty Level	Vector Space Correct Answer	Vector Space Incorrect Answer	All topics/subtopic Correct Answer	All topics/subtopic Incorrect Answer
Level 1	723	745	2250	2305
Level 2	337	438	1427	1651
Level 3	215	291	759	1078

4 Results and Discussions

To group students in different profiles and analyze their similarities and dissimilarities, the datasets were evaluated by the Fuzzy c-means algorithm.

For a comprehensive examination of student profiles through clustering, the number of correct and incorrect answers across three difficulty levels was assessed for each student. Outlier students were identified using the Z-score method using a cut-off of 2. A high Z-score (typically greater than a threshold, such as 2 or 3) indicates that the data element is far from the mean and can be considered an outlier [8], leading to their exclusion from the dataset. This refinement process left 131 students for Vector Space analysis and 331 for the dataset of all topics/subtopics. It is important to mention that the outlier analysis was conducted independently for each dataset.

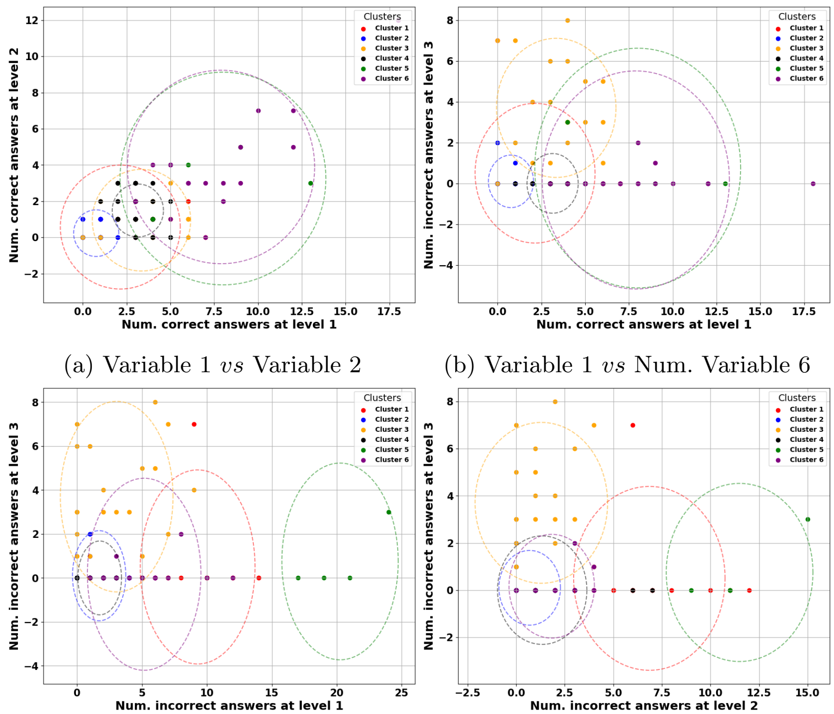
To execute the clustering analysis, six variables were delineated, each detailing the number of questions answered by a student across the difficulty levels and type of answer, as defined below.

- Variable 1: number of correct answers in questions of level 1.
- Variable 2: number of correct answers in questions of level 2.
- Variable 3: number of correct answers in questions of level 3.
- Variable 4: number of incorrect answers in questions of level 1.
- Variable 5: number of incorrect answers in questions of level 2.
- Variable 6: number of incorrect answers in questions of level 3.

In the FCM, all six variables were considered simultaneously for both datasets. The Elbow method was utilized [9], indicating 6 clusters for both datasets. Another parameter that must be defined in the FCM is the fuzzy partition exponent m, considered $m = 2$, as suggested by [1]. The obtained results for each dataset are presented below.

4.1 Vector Space results

In contrast to the crisp clustering method, which rigidly assigns each element to a single cluster, FCM adopts a more flexible approach. In fuzzy clustering, an element of the dataset can belong simultaneously to all clusters but with different degrees of membership (association). This means that, for the given problem, each student can be associated with a certain degree of membership to all clusters present in the dataset, reflecting the inherent uncertainty or ambiguity in data categorization. The FCM was applied to the Vector Space dataset, allowing for the identification of nuances in the distribution of the students by adopting this more flexible approach. Figure 1 presents the FCM results through four two-dimensional graphs, in which each dot represents a students. However, it is crucial to note that the dimension of this problem is six order, meaning that some elements of the set may lie outside the bounds of the ellipses, depending on the axis visualization. As can be observed, the distribution of the students in the space results in a fuzzy environment, rendering the fuzzy methodology suitable for analysis.



(a) Variable 1 *vs* Variable 2 (b) Variable 1 *vs* Num. Variable 6

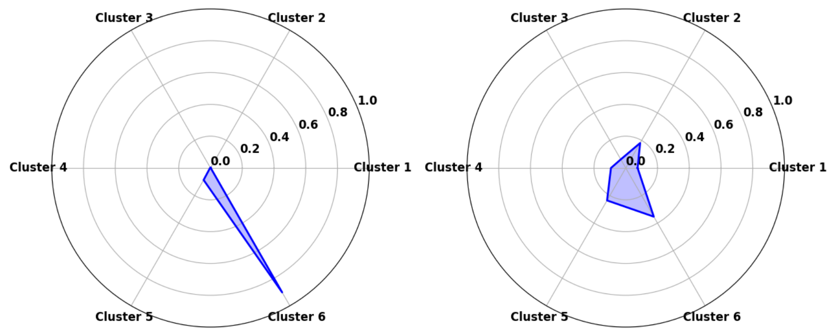
(c) Variable 4 *vs* Variable 6 (d) Variable 5 *vs* Variable 6

Fig. 1: Fuzzy c-means results of Vector Space dataset.

As mentioned, in the fuzzy clustering approach, students exhibit degrees of association with all considered clusters, ranging from subtle associations to those that stand out more prominently. To better illustrate this concept, consider the case of students with numbers 1071 and 1308, illustrated in Figure 2.

The student 1071 demonstrates a predominant association with cluster 6, with a degree of membership practically imperceptible in the other clusters. On the other hand, student 1308, although showing a higher degree of membership in cluster 6, also exhibits a significant association with clusters 2 and 5. In the crisp approach, both students would be grouped into the same cluster (cluster 6), implying similar performance in the vector space subtopic. This contrast highlights the flexibility of the fuzzy approach, allowing for a more detailed analysis and precise associations, revealing nuances that may go unnoticed in the crisp methodology. Table 2 shows the degrees of membership of 15 out of the 121 students (randomly selected) that compose the Vector Space dataset. Whereas Table 3 presents the number of questions answered by the students, according to the type of answer and the level of the question.

As can be observed, some students present high membership degrees in certain clusters, which poses no issues for using a clustering crisp method, as they



(a) Student 1071 (b) Student 1308

Fig. 2: Student's 1071 and 1308 degree of membership.

are strongly associated with a specific cluster, with a membership degree exceeding 0.9000. Student 1071 (cluster 6) and student 1147 (cluster 1) exemplify this situation. However, in most cases, this is not the scenario. Most students have representative membership degrees with similar values across multiple clusters. For instance, student 890 (clusters 2 and 4) and student 1087 (clusters 1 and 4). Cases like these present challenges for defining a single cluster, especially the case of student 1087, where the lowest membership degree is in cluster 3 and the highest in cluster 4 with a difference of only 0.1257.

Given that six axes and six clusters are considered, it is challenging to determine in which cluster the students with excellent or poor performance are located. This is because, when considering the number of questions answered, the results also reflect the students' engagement, i.e., who answers more or fewer questions at each of the three levels. Therefore, the objective of this work is not to identify the degree of excellence of the students but rather to determine how many categories of students exist (number of clusters) and to group them according to their main characteristics, such as the number of correct and incorrect answers and the number of responses given, which is also related to the student's engagement with the platform.

In general, by combining the results from Tables 2 and 3, it can be observed that students with significant membership in cluster 1, despite having answered many questions, have a high number of incorrect answers, such as students 1147 and 1152. On the other hand, students with representative membership in cluster 2 also exhibit a high quantity of incorrect answers and answer fewer questions on the platform, like students 966, 1157, and 1352. As for clusters 3 and 4, it is possible to find students with more correct answers, with those in cluster 3 showing higher engagement, such as students 633 and 1284. Lastly, students in clusters 5 and 6 are those who answer fewer questions and have a low number of correct answers, like students 1163, 1308, and 1334.

Table 2: Student's degree of membership for each cluster for the Vector Space

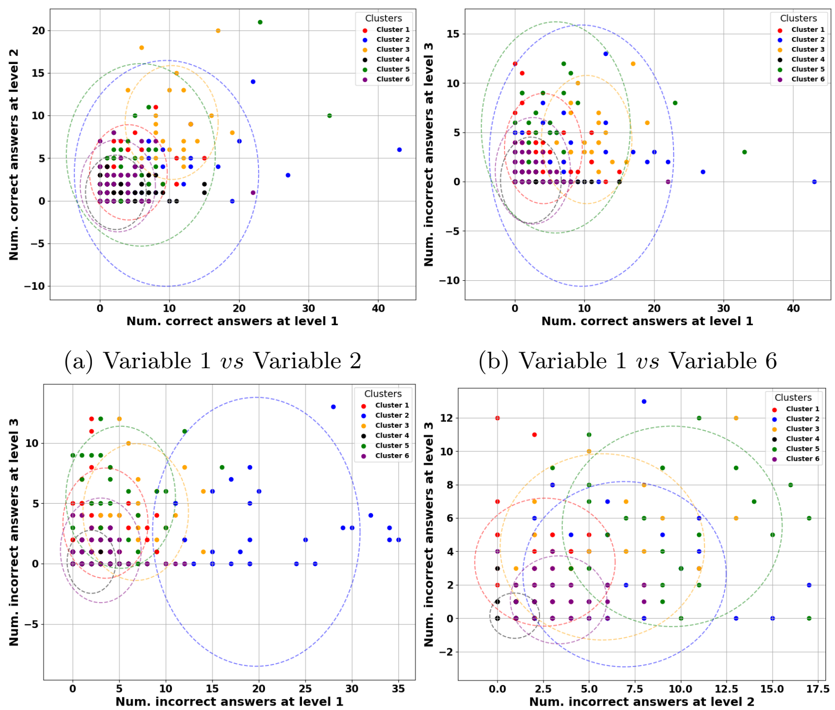
Student ID	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
411	0.00365	0.01047	0.00717	0.00866	0.75801	0.21205
633	0.08048	0.08299	0.52623	0.07789	0.11551	0.16872
890	0.09999	0.21838	0.12063	0.28683	0.13073	0.14342
966	0.02153	0.79025	0.01625	0.05674	0.05856	0.05663
1071	0.00133	0.00382	0.00226	0.00306	0.08688	0.90264
1080	0.23526	0.17162	0.07764	0.24819	0.11556	0.15173
1087	0.20660	0.14932	0.11272	0.23842	0.13194	0.16100
1147	0.92004	0.02190	0.01290	0.01412	0.01433	0.01668
1157	0.02551	078344	0.01793	0.05771	0.05825	0.05713
1163	0.00940	0.02740	0.01511	0.02015	0.41084	0.51710
1252	0.73507	00.6707	0.04379	0.04833	0.04812	0.05759
1284	0.09791	0.09511	0.48171	0.08197	0.12163	0.12164
1308	0.07462	0.18113	0.06250	0.09327	0.23543	0.35304
1334	0.07739	0.22351	0.06537	0.08563	0.24407	0.30402
1352	0.02888	0.59150	0.02322	0.15121	0.09954	0.10565

Table 3: Questions answered in vector space subtopic

Student ID	Level 1 Correct	Level 2 Correct	Level 3 Correct	Level 1 Incorrect	Level 2 Incorrect	Level 3 Incorrect
411	1	1	0	1	0	0
633	4	2	8	0	1	6
890	8	2	5	8	3	2
966	0	0	0	7	6	0
1071	3	1	0	2	1	0
1080	12	7	0	12	4	0
1087	18	12	0	10	2	0
1147	9	5	0	19	11	0
1157	0	0	0	7	7	0
1163	1	2	0	3	1	0
1252	13	3	0	17	9	0
1284	4	0	6	6	2	8
1308	3	3	0	3	6	0
1334	3	1	0	3	7	0
1352	2	1	0	8	3	0

4.2 All topics/subtopics results

Extending the methodology to a larger dataset, comprising all students who answered questions on the MathE platform in any available topics or subtopics, the Fuzzy c-means consistently produces results similar to those of the Vector Space dataset. Figure 3 illustrates the combination of some axes, in which each dot represents a students of the platform.



(a) Variable 1 *vs* Variable 2 (b) Variable 1 *vs* Variable 6

(c) Variable 4 *vs* Variable 6 (d) Variable 5 *vs* Variable 6

Fig. 3: Fuzzy c-means all topics.

Once again, there is a very diffuse dataset that is appropriate for using fuzzy methods. Table 4 presents the membership degrees of 15 out of 297 students (randomly selected) who compose the dataset of all topics/subtopics. Table 5 describes the number of questions answered and the type of answers the 15 selected students provided. When employing a larger amount of data, the nebulous nature of the data becomes even more apparent. As observed in Table 4, only student 1156 exhibited a membership degree above 0.8000 in one of the clusters; all other students have membership degrees quite distributed across more than one cluster. This outcome was expected since we are considering all the information from the topics and subtopics that compose the MathE platform.

The students who exhibit a stronger association with cluster 1 (students 174 and 422) achieved more correct answers at level 1 and a considerable amount of errors at level 2. However, at level 3, they had similar incorrect and correct answers. They answered more questions at level 2 than at level 1. In cluster 2, (students 967 and 1334), they answered several questions at all levels, but more at level 1.

Table 4: Student's degree for membership considering all topics/subtopics

Student ID	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
174	0.63634	0.12950	0.05421	0.07115	0.06074	0.04806
403	0.01715	0.02952	0.03266	0.61458	0.22288	0.08321
422	0.70104	0.09741	0.05950	0.05784	0.04689	0.03731
689	0.02416	0.03963	0.05266	0.63975	0.16854	0.07527
890	0.19605	0.14817	0.37489	0.11116	0.09333	0.07638
965	0.01762	0.03327	0.02687	0.14502	0.29297	0.48426
967	0.06969	0.41858	0.10046	0.17449	0.12977	0.10698
1071	0.00813	0.01417	0.01081	0.05842	0.15344	0.75502
1077	0.21646	0.10780	0.32747	0.14306	0.11837	0.08682
1080	0.19701	0.11847	0.25906	0.14951	0.14260	0.13332
1087	0.20431	0.13956	0.23542	0.14829	0.14159	0.13080
1156	0.00636	0.00833	0.01056	0.04485	0.11853	0.81134
1163	0.01241	0.02498	0.01817	0.11756	0.32592	0.50096
1334	0.13159	0.31981	0.07976	0.21164	0.15659	0.10061
1352	0.06204	0.14864	0.07600	0.31035	0.23834	0.16463

Table 5: Questions answered considering all topics/subtopics

Student ID	Level 1 Correct	Level 2 Correct	Level 3 Correct	Level 1 Incorrect	Level 2 Incorrect	Level 3 Incorrect
174	7	2	3	3	10	3
403	1	4	2	3	4	2
422	5	3	5	3	11	5
689	2	5	2	3	4	3
890	13	4	5	29	10	3
965	7	3	0	4	0	0
967	8	5	8	7	2	3
1071	4	0	0	3	0	0
1077	4	0	0	15	11	6
1080	27	3	1	18	1	1
1087	43	6	0	15	6	0
1156	3	0	0	2	0	0
1163	6	2	0	5	1	0
1334	5	1	0	11	6	5
1352	3	3	3	9	2	1

In cluster 3, there are students who answered the most questions on the platform, especially at level 1, students 890, 1077, 1080, and 1087. Therefore, these students are the most active, especially the students 1080 and 1087, which showed many correct answers in level 1. The students with stronger association in cluster 4 (students 689 and 1352) answered fewer questions and had a similar number of correct and incorrect answers. The students with the highest association in cluster 5 are also strongly associated with cluster 6 (students 965 and 1163). These students respond to a few questions, mostly at level 1, and have moderate accuracy compared to the number of errors. Overall, students in different clusters show varying engagement patterns, performance, and accuracy across different levels of questions, resulting in a complex space more suitable for employing fuzzy methodologies. However, it is essential to apply feature selection techniques to achieve a more effective interpretation of the solutions.

5 Conclusion

This paper uses the Fuzzy c-means algorithm to analyze the categories in which students, who use the MathE platform, can be grouped. The Z-score method [8] was employed to refine the data and eliminate the students considered outliers. Additionally, the Elbow method [9] was used to determine the optimal number of clusters. Afterward, the Fuzzy c-means algorithm was employed to categorize the datasets, considering six variables that answer questions from Vector Space subtopic and a second one that encompasses all topics and subtopics of the platform. Considering this, the goal was to explore a fuzzy clustering method to categorize the students according to their behavior on the platform.

The Fuzzy c-means algorithm is relevant in education, where students' skills and knowledge do not fit into rigid categorizations. Fuzzy clustering assigns multiple memberships to students, enabling a comprehensive approach to data analysis. This is crucial for identifying students with various group characteristics and enhancing understanding of intersections between knowledge areas and skills. Such an approach enriches analyses, promoting a nuanced understanding of learning patterns and academic performance.

Each person possesses unique strengths and weaknesses across various scientific domains. Success in education and career often depends on the adeptness at using one's strengths and addressing areas of weakness [11]. The obtained results demonstrate the effectiveness of Fuzzy c-means in an educational context, and also highlight the importance of more flexible approaches in the face of the intrinsic complexity of academic performance data. To further advance this study, it is intended to explore feature importance techniques and dimensionality reduction, allowing the extraction of more significant patterns to enrich the project development in terms of personalization and system recommendation. Moreover, the reduction of variables also provides a more effective evaluation of students according to their level of proficiency in each difficulty level of the platform's topics and subtopics, as well as student engagement while the platform resources users. Soon, there will be a focus on developing tools tailored to support individualized learning processes.

References

1. Aggarwal, C.C., Reddy, C.K.: Data Custering Algorithms and Applications. Taylor & Francis Group. CRC Press (2013)

2. Alhawiti, N.M.: The influence of active learning on the development of learner capabilities in the college of applied medical sciences: Mixed-methods study. Advances in medical education and practice **14**, 87–99 (2023)
3. Azevedo, B.F., Fernandes, F.P., Pacheco, M.F., Pereira, A.I.: Dataset for assessing mathematics learning in higher education. `https://doi.org/10.34620/dadosipb/PW3OWY` (2024)
4. Azevedo, B.F., Leite, G., Pacheco, M.F., Fernandes, F.P., Rocha, A.M.A.C., Pereira, A.I.: Multi-objective clustering algorithm applied to the mathe categorization problem. Submitted to Information Systems Frontiers
5. Azevedo, B.F., Pacheco, M.F., Fernandes, F.P., Pereira, A.I.: Dataset of mathematics learning and assessment of higher education students using the mathe platform. Data in Brief **53**, 110236 (2024)
6. Azevedo, B.F., Rocha, A.M.A.C., Pereira, A.I.: Hybrid approaches to optimization and machine learning methods: a systematic literature review. Journal of Machine Learning (2024)
7. Azevedo, B.F., de Souza, R.M., Pacheco, M.F., Fernandes, F.P., Pereira, A.I.: Application of pattern recognition techniques for mathe questions difficulty level definition. In: Pereira, A.I.e. (ed.) Optimization, Learning Algorithms and Applications. pp. 300–315. Springer Nature Switzerland, Cham (2024)
8. Carpenter, J.R., Hoffmann, T.P.: Statistics for Data Science: A Comprehensive Introduction. Wiley (2019)
9. Delgado, H., Anguera, X., Fredouille, C., Serrano, J.: Novel clustering selection criterion for fast binary key speaker diarization. In: 16th Annual Conference of the International Speech Communication Association (NTERSPEECH 2015) (2015)
10. Deslauriers, L., McCarty, L.S., Miller, K., Callaghan, K., Kestin, G.: Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. Proceedings of the National Academy of Sciences **116**(39) (2019)
11. Gardner, H.: Frames of mind: The theory of multiple intelligences. No. 3, New York: Basic Books, USA (2011)
12. Gu, X., Tan, Y., Wu, X., Cai, Z., Lai, Q., Cheng, M., Zhao, Q.: Active versus passive strategy in online creativity training: How to best promote creativity of students with different cognitive styles? Thinking Skills and Creativity **44** (2022)
13. Han, H.: Fuzzy clustering algorithm for university students' psychological fitness and performance detection. Heliyon **9**(8), e18550 (2023)
14. Hwang, G.J.: Definition, framework and research issues of smart learning environments - a context-aware ubiquitous learning perspective. Smart Learning Environments **1**(4) (2014)
15. Inyang, U.G., Joshua, E.E.: Fuzzy clustering of students' data repository for at-risks students identification and monitoring. Comput. Inf. Sci. **6**, 37–50 (2013)
16. Zhi, L.: Research on the influence of mental health on college students' employment based on fuzzy clustering techniques. Soft Computing **27**, 19095–19111 (2023)