University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.17 No.4 Dec 2022

Students Groups Detection in Online Examinations Using K-Means Clustering

Ahmad Al Musawi[1], Ali Hussein Lazem[2] , Mustafa Asaad Hasan[3]

[1]Department of Information Technology, University of Thi-Qar, Thi Qar, Iraq .

[2,3]University of Thi-Qar,

[1]almusawiaf@utq.edu.iq, [2]alazem@utq.edu.iq, [3]mustafa.alkhafaji@utq.edu.iq.

https://doi.org/10.32792/utq/utj/vol17/4

ABSTRACT

Schools and universities have been adversely affected by the widespread distribution of COVID-19 and related quarantine around the worlds. As a result of this distribution, most of these institutions have deployed online teaching platforms as an alternative to students' physical attendance. However, the usages of recent online technologies have provided extra communication channels in addition to e-learning media. Data availability and accessibility have made it possible to conduct online searches. Studying the students' performance in the online examination is conducted to determine the degree of similarity and groups of students who shared similar behavior. The K-Means clustering model has been implemented on the tf-idf representation of the retrieved online corpus. The study concludes that students fall into five distinguished groups (i.e. small communities) based on similarity in performance of sharing the same significant content over the different courses. A larger corpus (document collection) of the complete academic performance of students at different levels (as future work) would help refine more accurate groups of collaboration among students.

Keywords: K-Means Clustering, Online Examination, Students Group

1. INTRODUCTION

Due to the wide distribution of the COVID-19 pandemic in December 2019, most universities and schools around the world have converted to the implementation of online education to preserve the academic development of their students after repeated cessations. Online education has included a wide range of responsibilities like setting the curriculum, and schedules, sharing video lectures (lived or recorded), adding tests and assignments, and so on. Several online applications provided a comprehensive list of services to fulfill academic requirements, like Google

Classroom, Canvas, Blackboard Learn, Moodle, Edmodo, etc. Such technologies have been widely distributed [1].

With the implementation of these services, students have been able to share advanced solutions and methods while working in an environment that is conducive to sophisticated learning. However, while communication in e-learning platforms is observed and administered by the course instructor, several online social media widely exist on the web alongside. Social media like Facebook, Telegram, Instagram, etc. have been used by a wide range of students as a medium for sharing related news and university requirements in a matter of seconds. These media have been aligned with the e-learning application system to distribute information among connected students outside the scope of the instructor, as a discussion forum and debate.

It is very critical to have an insight into the orientation of students' groups that share information of either answers or insights of answers. One major disadvantage of using online e-learning applications is cheating. The availability of massive amounts of information on the world wide web (WWW) has made the accessibility to information a matter of several clicks. Students can reformat the knowledge received from internet sites into an accepted solution. Therefore, students may result in having a set of varied different answers to the same question. As a result, we may end up having groups of trust that share the approved knowledge among students. A student (who agrees to cheat) may accept a specific format of an answer that has a higher possibility of being shared and deployed as the final answers.

Recent studies have focused more on the behavioral dynamics of social actors [2, 3]. In other words, the social behavior of humans and animals tends to participate in groups for survival. In this study, we aimed to analyze the behavior of students and their tendency to affiliate into groups of trust. A trust group refers to a set of actors (students) that believe in shared knowledge. It is very important to mention that we did not interpret the similarity in solution as cheating or not. However, the following algorithm only measures the degree of similarity among the given observations. The targeted aim is to conduct the groups of students that mostly share the same performance over categorized answers.

A wide range of models and algorithms are being used for analyzing educational data. Baker et al. [4] specified five basic technical methods for analyzing online educational platforms: clustering, prediction, discovery models, distillation for human judgment, and relationship mining. However, to approach the groups of students, a clustering model has been implemented herein. Clustering algorithms are unsupervised machine learning models that combine entities with similar features into one group. There are many clustering models with different parameters. Most of these models differ in the ways of measuring the similarity of different objects and the way of deciding which two objects belong to what group. Features of the objects can be a collection of values of different types (numerical, string, etc.). The similarity is a central concept in clustering, as it refers to the degree of likelihood between two objects. If two objects share the same values for the same features, then these objects are similar, and their similarity equals one.

Group detection plays a critical role in understanding the connectivity shared by participants and can improve the level of students' education acceptability and interaction. Educational groups are built by either direct assignment which is mostly made by the teacher or by students' self-allocation. Moreover, a parallel self-association of groups may exist as a result of online examinations or students' self-teaching opportunities. Several psychological and social factors contribute to the generation of new relationships among students. Trust and confidence, similar social and conceptual affiliation, and level of mind-compatibility represent vital influencers for making a connection among students.

The impact of these two categories on the results of the assignments can help teachers and the education administration in setting constructive rules for group establishment. Furthermore, these rules can be embedded within artificial models that read the performance of given students and learn which group structure may result in better performance for the majority of students.

2. RESEARCH METHOD

The objectives of increasing the level of student engagement and academic performance have been the focus of sophisticated ongoing research [5, 6, 7, 8]. Two basic categories of research have been conducted in the literature that serves these objectives. The first category is based on understanding the student's behavior during the course period. It has been concluded that collaborative learning has shown a high performance and increased the degree of student engagement in the course[9]. Student engagement in online learning platforms is considered one of the ongoing research that aims at understanding the contribution of various tools and collaborative approaches for better student academic performance. Several factors measure student engagement in online learning platforms like biometric-based registration, online attendance, assignment submission, group studies, and participation in different academic activities. A crucial research question was raised on quantifying the significance of the individuality of a student or his/her participation in groups in achieving the purpose of the academic course. Moreover, understanding the patterns of association among these different factors helps in determining the standards for the systematic learning process.

The second category is based on understanding the overall performance of students. In [10], a study on 3268 students has shown that behavioral engagement (as a determination in learning) and a sense of affiliation to groups have a major role in predicting the performance of the students. Moreover, students have fallen into three groups of performance, based on a study on 75 students that have been conducted to understand the behavior of students in their assignment submission and its relation to their success, [11]. In [12], Two-phases hierarchical algorithms have been used to group students based on their learning preferences (learning path).

The k-Means clustering model has been used for grouping students in Virtual Learning Environment (VLE). Five categories (groups) of students have been detected: expert, good, regular, bad, and criticism answers, [13].

3. METHODS

Several methods are used in measuring the similarity between a given query and document collection (corpus) in any given information retrieval system, such as the Boolean model, Bag of Word model (BoW), vector space model, or term weights-based model of TF-IDF. In the boolean model, the words of a query are assigned a boolean value based on their existence in the corpus or not. Likewise, the Bag of Words model checks the multiple occurrences of the same elements (words) in both the query and document collection. In the vector-space model, a word is assigned a continued value of its occurrence in each document. All of the previous model do not consider the relative weight of a word in comparison to its existence in different documents. However, the most significant model that assigned a normalized and relative weight to words is the TF-IDF.

3.1. **TF-IDF**

To score a word in a given solution, the TF-IDF measure has been implemented, [14]. It is a widely used method for information retrieval and document search. It scores word relevance to a document (solution) in a set of documents. TF-IDF stands for Term Frequency- Inverse Document Frequency. TF-IDF measures the significance of a word to one document in comparison to its appearance in all other documents. Term frequency (TF) would count the number of appearances of a word in one document. The inverse document frequency (IDF) will score a word in the range [0-1], depending on its common or rare existence in the documents. As the IDF of a word is closer to 0, then the word is most common and vice versa. The IDF score is calculated by taking the log as a result of dividing the number of solutions by the number of solutions that contain the word, see Eq. 1.

$$tf(t,d) = freq(t,d) = \frac{number\ of\ times\ t\ appears\ in\ d}{total\ number\ of\ terms\ in\ d}$$

$$idf(t,D) = log\ log\ \left(\frac{N}{count(d \in D: t \in d)}\right)$$

$$tf - idf(t,d,D) = tf(t,d) * idf(t,D)$$

Where t, d, D, N to refer to the term, current document, collection of documents and number of documents, respectively. The implementation of the TF-IDF will result in an array where columns are equal to the number of terms in the total corpus, and rows are equal to the number of documents. Each element in the array will represent the $tf - idf$ value of the given (term, document).

3.2. **K-Means Clustering**

K-Means clustering model is one of the most unsupervised machine learning models used to partition the observations dataset into k groups (mostly called clusters), [15]. k refers to the number of clusters that are to be specified by the model or the analyst. Each group/cluster contains a set of observations that share maximum similarity to other observations in other groups. The ideas behind the clustering model are as follows:

1. Specify a value of k, the number of clusters.
2. Select a k random item within the observation as a cluster center, mostly referred to as centroid.
3. Assign each data item to the closest centroid based on Euclidean distance.
4. Update the cluster's centroid by considering the mean values of all data items in each cluster. This operation is done by taking the average of each attribute and comparing it with the centroid.
5. Repeat 3-4 until centroids changes decrease to a specific range or until a maximum number of iterations is reached.

The previous algorithm is considered a widely popular one. However, several other models are much faster and with different approaches, which details are out of the scope of this paper.

3.3. **Implementation**

The dataset being used for this study has collected the answers of the students for the final first attempt examination, in the university of Thi Qar, College of Computer Science and Mathematics, Department of Computer science. The examinations were done using Google Classroom for three topics: machine learning (ML), computer graphics (CG), and data structures (DS) and downloaded as an xlsx file. Table. 1 shows the dataset statistics. To identify the groups of students that performed similarly, we conducted three levels of analysis: data cleaning, answers clustering, and student performance clustering, see Figure. 1. Data cleaning is a primary step to prepare the data for the next analysis steps. The input data contains detailed information about each student, like full name, email, time of submission, stage, type of study, and so on. Most of this information is structurally irrelevant observations and has been removed. The students' private information has been replaced with a specific identifier for later linkage. Also, rows of data that contain links to files uploaded have been removed as well.

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.17 No.4 Dec 2022

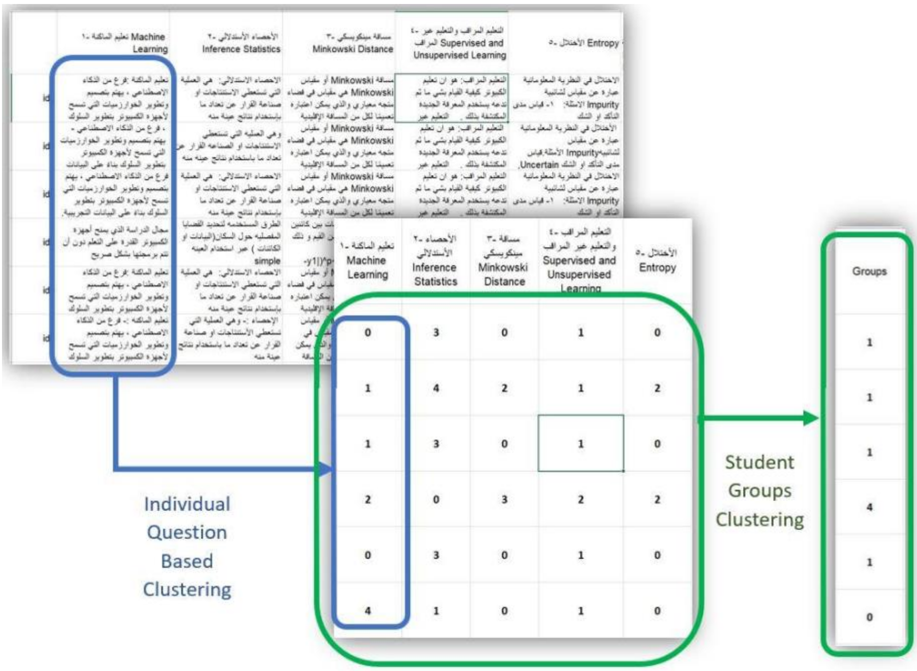


Figure 1. Workflow of detecting groups of students in the online examination.

Table 1. Dataset Statistics

Stage	Material	Participants	Questions
3rd stage	Machine Learning (ML)	104	5
3rd stage	Computer Graphics (CG)	100	3
2nd stage	Data Structure (DS)	49	6

The K-Means clustering model has been used to divide the answers into groups of solutions. Each group would represent answers with similar content. Herein, we collect the most similar answers into clusters for simplifying the distribution of students in the student clustering. As the answer words used for one question are different from those words used for others, we only consider each question separately for the clustering. For example, Q1 in the CG examination has resulted in 5 groups of similar answers and so on for the rest of the questions, see Table. 2.

Table 2. the number of clusters for each question, based on its Elbow values.

Material	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**
Machine Learning (ML)	6	5	6	4	3	-
Computer Graphics (CG)	7	4	6	-	-	-
Data Structure (DS)	3	6	7	7	5	4

To better classify the types of answers given by various students, a K-Means clustering model used. However, one of the most important factors in interpreting the answers is to match the validity of the student's answer with the approved answer. This can help in detecting the rightest solution among the solutions of several students. Herein, natural language processing NLP models are needed to break up the two given texts into their conceptual, syntactic, and symbolic structures and implement searching and matching of the existence of specific sets of terminologies. The comparison of two terminologies is a challenging task in NLP as one sentence can be written in several forms and yet all refer to the same meaning. This will add an extra level of complication to our mission. In this study, however, we only care about the similarity among students' performance rather than evaluating the solutions. The determination of students' groups may end up detecting cheating groups as well, although adding NLP models would lead to better determination. Consequently, an abstract comparison of the two texts has been implemented in this study. The comparison simply searches for the maximum similarity between two sequences of words based on their TF-IDF values.

After specifying the clusters of each separated question. We target to cluster students based on their performance. The solution of each student has been replaced with its corresponding answer cluster. Now, another level of clustering is implemented using K-Means for student overall performance. The result of the last clustering ends in grouping students into clusters of similar performance, i.e. students that mostly share similar answers.

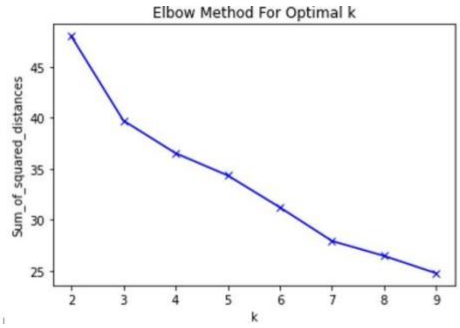


Figure 2. Effects of selecting different switching under dynamic conditions

4. RESULTS AND DISCUSSION

Each question in each material has resulted in a different number of answer clusters, see Table. 2. The k values for each question have been calculated using the Elbow method for optimal k, which gives k that provides the minimum total within-cluster variation (or sum of the square, wss). A simulation is conducted to measure the minimum wss for each question by selecting the bend that appears in the plot, see Figure. 2.

Different questions resulted in various clusters. Table. 3 shows the percentages of students' participation in groups for the given question. We can see that a large percentage of students participate in one major group in each question, mostly falling into 51\% on average. This result comes up because most of the students' answers use the same terminologies for the answers. The new matrix used in the next step is created by assigning the groups of answers for each question instead of string-based answers.

A clustering model is implemented based on answers groups on each question that resulted in groups of students' performances in the online examination. The same K-Means method was implemented for each subject and we got 5 clusters for each subject. The percentages have been collected for each cluster for ML, CG, and DS materials, see Table.4. Cross-material averages show that students have participated in the dominant answers' clusters, sharing the same crowd performance. We are also illustrating the distribution of answer groups in each cluster, see Figures 3, 4, and 5.

Groups of answers are not related to different questions. However, answers' groups may appear over different students' clusters. By implementing the K-Means, we have found the distribution of students into groups based on the category achieved on answers clustering.

Three different materials examinations have revealed the structure of students into groups. About 34% of the students share the same performance in terms of using a similar set of relevant words in answers. In ML examination, we can see that the number of the groups within one cluster are distributed over three groups. Most of the students share the same answer' groups in each cluster. ML examination has been set up in such a way of forcing the student to explain in detail the content of the mathematical measures. For example, Q6 has resulted in different answer groups over the five student clusters. In CG and DS examinations, a specific formal definition is required which reduces the number of answers' groups lower than those in ML examination.

Table 3. Sorted percentage of students in each question group. M: Material, Q: Question, G: Group.

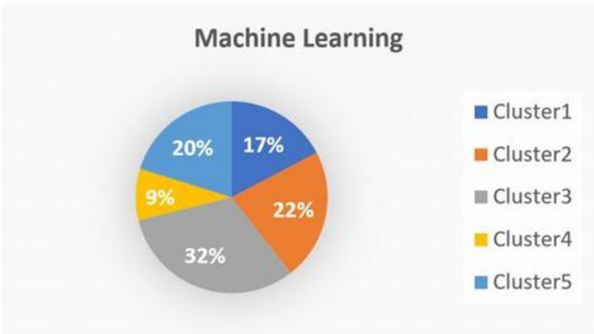
M	Q	G1	G2	G3	G4	G5	G6	G7
Machine Learning (ML)	1	0.45	0.14	0.14	0.1	0.1	0.07	0
	2	0.39	0.21	0.2	0.12	0.09	0	0
	3	0.47	0.14	0.13	0.11	0.1	0.07	0
	4	0.69	0.12	0.11	0.09	0	0	0
	5	0.4	0.32	0.28	0	0	0	0
	1	0.47	0.16	0.11	0.08	0.07	0.07	0.04

Computer Graphics (CG)	2	0.41	0.24	0.23	0.12	0	0	0
	3	0.38	0.37	0.12	0.05	0.04	0.04	0
Data Structure (DS)	1	0.76	0.16	0.08	0	0	0	0
	2	0.47	0.39	0.08	0.02	0.02	0.02	0
	3	0.82	0.06	0.04	0.02	0.02	0.02	0.02
	4	0.39	0.27	0.1	0.08	0.06	0.06	0.04
	5	0.61	0.12	0.1	0.08	0.08	0	0
	6	0.41	0.29	0.2	0.1	0	0	0

Table 4. Sorted Percentages of final students' clusters.

Cluster	ML	CG	DS	Average
1	32%	30%	39%	34%
2	22%	25%	27%	25%
3	20%	17%	12%	16%

4	17%	17%	12%	15%
5	9%	11%	10%	10%



University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.17 No.4 Dec 2022

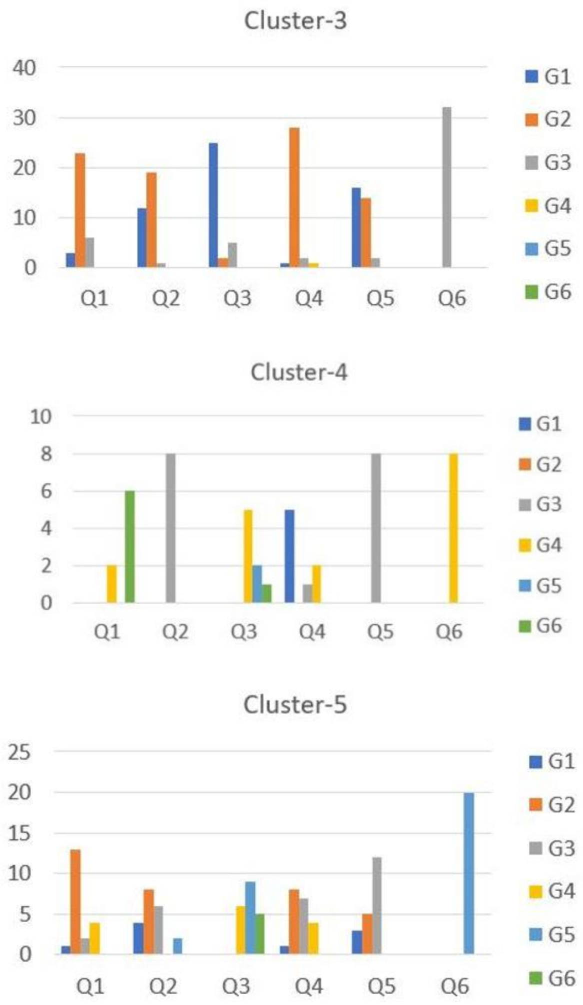
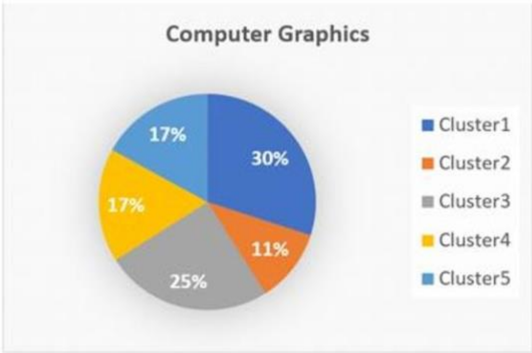
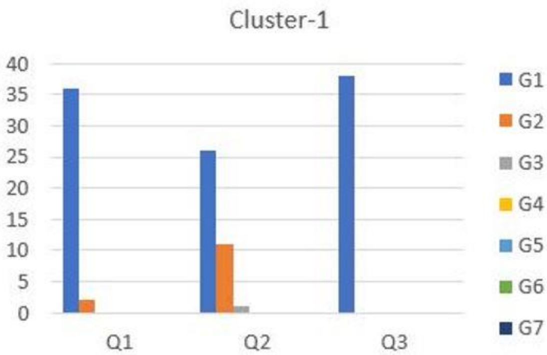


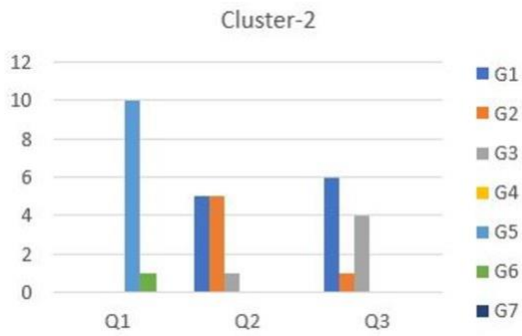
Figure 3. Effects of selecting different switching under dynamic condition (Machine Learning).



Computer Graphics
30%
11%
25%
17%
17%
Cluster1
Cluster2
Cluster3
Cluster4
Cluster5



Cluster-1
40
35
30
25
20
15
10
5
0
Q1
Q2
Q3
G1
G2
G3
G4
G5
G6
G7



Cluster-2
12
10
8
6
4
2
0
Q1
Q2
Q3
G1
G2
G3
G4
G5
G6
G7

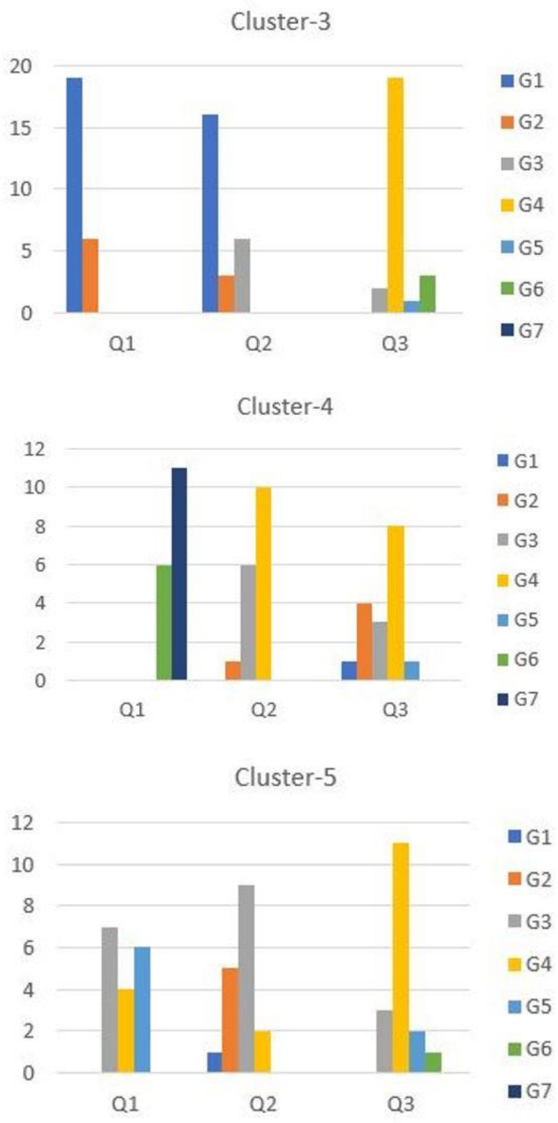


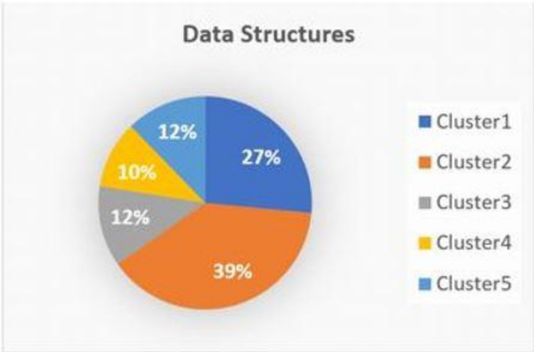
Figure 4. Effects of selecting different switching under dynamic condition

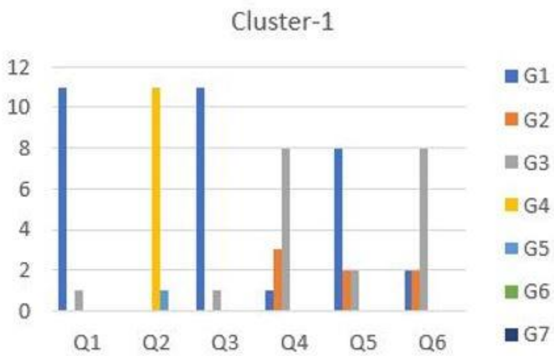
(Computer Graphics)

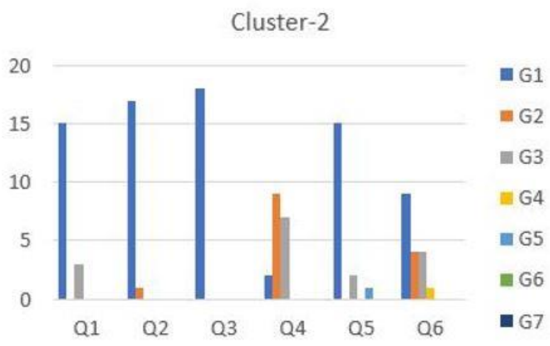
University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.17 No.4 Dec 2022







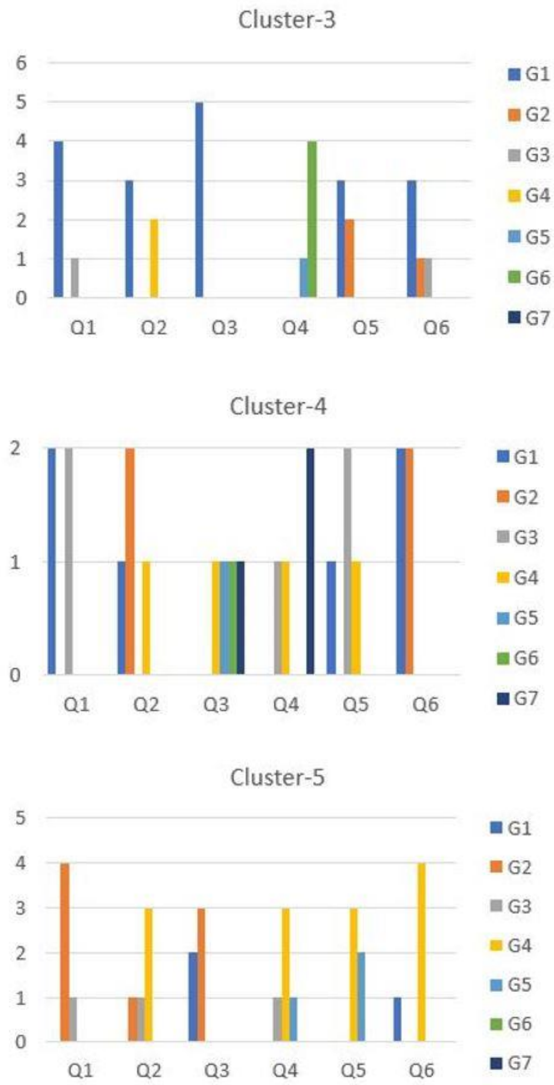


Figure 5. Effects of selecting different switching under dynamic condition (Data Structures)

5. CONCLUSION

This paper proposes implementing the K-Means clustering algorithm that groups a set of students with similar exam performance into coherent interacting units (clusters). We have found that most of the students' performance falls into five distinct clusters in each examination. Student groups are easily detected using a two levels clustering paradigm.

This kind of research can be extended to understand the relationship among students based on additional factors like gender, age, address, and so on. Further analysis of students interaction would utilize most of the implemented models herein, however, an NLP-based approach will further extend the conceptual based analysis of the used terminology amongs the groups and associate them with groups social and scientific factors. Moreover, a comprehensive analysis of the complete academic performance of severl stages could improve the overall understanding of dynamics of interaction in the given academic environment.

References

[1] I. Elaine Allen and Jeff Seaman.Changing course: Ten years of tracking online education in the United States.ERIC, 2013.

[2] Ren ́e Veenstra and Christian Steglich. Actor-based model for network and behavior dynamics. 2012.

[3] Tom AB Snijders. The statistical evaluation of social network dynamics.Sociological methodology, 31(1):361–395, 2001.

[4] Ryan SJD Baker, Sujith Gowda, and Albert Corbett. Au-tomatically detecting a student’s preparation for futurelearning: Help use is key. InEducational Data Mining2011, 2010.

[5] Shaobo Huang and Ning Fang. Predicting student aca-demic performance in an engineering dynamics course:A comparison of four types of predictive mathematicalmodels.Computers & Education, 61:133–145, 2013.

[6] Louis Cohen, Lawrence Manion, and Keith Morrison.Research methods in education. routledge, 2013.

[7] Sarah Ransdell. Predicting college success: The impor-tance of ability and non-cognitive variables.International Journal of Educational Research, 35(4):357–364, 2001.

[8] Siu-Man Raymond Ting and R Man. Predicting academic success of first-year engineering students from standard-ized test scores and psychosocial variables.InternationalJournal of Engineering Education, 17(1):75–80, 2001.

[9] Maria J. Casuso-Holgado, Antonio I. Cuesta-Vargas,Noelia Moreno-Morales, Maria T. Labajos-Manzanares,Francisco J. Bar´on-L´opez, and Manuel Vega-Cuesta. The Association between academic engagement and achieve-ment in health sciences students.BMC Medical Educa-tion, 13(1):33, dec 2013.

[10] Jung-Sook Lee. The relationship between student engage-ment and academic performance: Is it a myth or reality?The Journal of Educational Research, 107(3):177–185,2014.

[11] G¨okhan AKC¸APINAR and Mehmet KOKOC¸. Analyzing The relationship between student's assignment submis-sion behaviors and course achievement through process mining analysis.T¨urk Bilgisayar ve Matematik E˘gitimiDergisi, 11(2):386–401, 2020.

[12] Danuta Zakrzewska. Using clustering technique for stu-dents' grouping in intelligent e-learning systems. InSym-podium of the Austrian HCI and Usability EngineeringGroup, pages 403–410. Springer, 2008.

[13] Alana M De Morais, Joseana MFR Araujo, and Evan-dro B Costa. Monitoring student performance using data clustering and predictive modelling. In2014 IEEE Fron-tiers in Education Conference (FIE) Proceedings, pages 1–8. IEEE, 2014.

[14] Juan Ramos et al. Using tf-idf to determine word rele-vance in document queries. InProceedings of the first instructional conference on machine learning, volume242, pages 133–142. New Jersey, USA, 2003.

[15] Gregory A Wilkin and Xiuzhen Huang. K-means clus-tering algorithms: Implementation and comparison. InSecond International Multi-Symposiums on Computer And Computational Sciences (IMSCCS 2007), pages 133–136. IEEE, 2007.