# Clustering Analysis in the Student Academic Activities on COVID-19 Pandemic in Mexico

G. Miranda-Piña[1(✉)], R. Alejo[1], E. Rendón[1], E. E. Granda-Gutíerrez[2], R. M. Valdovinos[3], and F. del Razo-López[1]

[1] Division of Postgraduate Studies and Research, National Institute of Technology of Mexico, (TecNM) Campus Toluca, Av. Tecnológico s/n, Agrícola Bellavista, 52149 Metepec, México
MM22280266@toluca.tecnm.mx

[2] UAEM University Center at Atlacomulco, Universidad Autónoma del Estado de México, Carretera Toluca-Atlacomulco Km. 60, 50450 Atlacomulco, México

[3] Faculty of Engineering, Universidad Autónoma del Estado de México, Cerro de Coatepec s/n, Ciudad Universitaria, 50100 Toluca, México

**Abstract.** The pandemic caused by the COVID-19 disease has affected all aspects of the life of the people in every region of the world. The academic activities at universities in Mexico have been particularly disturbed by two years of confinement; all activities were migrated to an online modality where improvised actions and prolonged isolation have implied a significant threat to the educational institutions. Amid this pandemic, some opportunities to use Artificial Intelligence tools for understanding the associated phenomena have been raised. In this sense, we use the K-means algorithm, a well-known unsupervised machine learning technique, to analyze the data obtained from questionaries applied to students in a Mexican university to understand their perception of how the confinement and online academic activities have affected their lives and their learning. Results indicate that the K-means algorithm has better results when the number of groups is bigger, leading to a lower error in the model. Also, the analysis helps to make evident that the lack of adequate computing equipment, internet connectivity, and suitable study spaces impact the quality of the education that students receive, causing other problems, including communication troubles with teachers and classmates, unproductive classes, and even accentuate psychological issues such as anxiety and depression.

**Keywords:** COVID-19 · *K*-means · Machine learning · Student academic activities

## 1 Introduction

Coronavirus disease (COVID-19) is an infectious respiratory illness caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). The pandemic caused by

This work has been partially supported under grants of project 11437.21-P from TecNM and 6364/2021SF from UAEMex.


D.-S. Huang et al. (Eds.): ICIC 2022, LNAI 13395, pp. 67–79, 2022.
https://doi.org/10.1007/978-3-031-13832-4_6

COVID-19 is a problem the world faces due to its high transmission and case fatality rate. COVID-19 was declared a public health emergency by the World Health Organization (WHO) just two months after the first outbreak was recorded in Wuhan, China at the end of 2019 [19]. The virus is transmitted from one person to another attacking the respiratory system of the host, exhibiting fever, muscle pain, dry cough, and shortness of breath (in severe cases). The symptoms appear in approximately 10–14 days, but although the carriers of the virus do not exhibit any symptoms, they can spread the virus without knowing (asymptomatic subjects) and its spread is quite accelerated [5, 20].

At the origin of the pandemic, there was no antiviral agent to treat the infection, or a vaccine [4], so prevention measures were the best strategy against COVID-19 [12]. Then, some measures included mandatory confinement in cities, temperature scanning to identify people with symptoms associated with this pathology, suspension of public transport, travel restrictions, and border closures, among others [2, 6, 8, 16, 18]. In the international context, the timely response of the government and public and private organizations was key to mitigating the possible risk of contagion. In the case of Mexico, on March 20, the General Health Council declared a health emergency derived from a rapid spread of positive cases. The initial strategy included voluntary and mandatory lockdown and considered the suspension of non-essential activities until April 30, 2020, because the magnitude of the pandemic was not yet known; however, as the pandemic evolved, the lockdown period lasted for two years. Although the pandemic has impacted almost all aspects of our daily lives, from social, financial, labor, and health, academic activities were especially affected. To mitigate the COVID-19 propagation in students, all universities in Mexico closed their facilities in March 2020, and face-to-face classes migrated to the online modality. An unprecedented disruption in the history of higher education in the country occurred, and the short, medium and long-term impacts are yet difficult to quantify.

The academic life of students and teachers has been affected in different ways; they were forced to enter an unplanned dynamic of online activities, which affected their daily lives and the continuity of their learning and mobility. The latter resulted in undesired changes in the behavior of the students. A lot of research on the behavior of COVID-19, from a medical point of view, has been performed [17], but also the impact of the pandemic on student behavior from the social and academic points of view is essential. Researchers from different fields of science (epidemiologists, health doctors, biochemists, etc.) are trying to learn more about the COVID-19, as well as its impact on the daily lives of citizens; then, there are currently many scientific papers that try to discover the behavior and evolution of patients [4]. Also, scientists have used artificial intelligence (AI), Big Data, and machine learning (ML) techniques to find trends in the evolution of COVID-19 [12], considering the large amount of data generated in hospitals and medical centers. These data are of different natures, such as demographics, previous illnesses, eating habits, and type of activity, among others [10].

AI techniques provide valuable insights and solutions for real-world problems based on data that otherwise are incomprehensible to humans. Clustering is a powerful machine learning approach widely used to analyze data. This is considered an unsupervised machine learning technique that uses unlabeled data to identify different trends in the data by creating groups or categories, even to allow to know what data features are of significant importance in the dataset. In this sense, we present a machine learning study based on a clustering approach to analyze the behavior of a group of students at a Mexican university with the aim of understanding how the COVID-19 pandemic has affected their academic life.

## 2 Related Work

Relevant literature based on the topic of inquiry is presented in this section, including statistics, machine learning, and clustering.

Campedelli et al. [3] present a study about the temporal distribution of civil disorders as a result of the coronavirus pandemic in 2020. Specifically, the temporal clustering and self-excitability of events were discussed. The authors proposed K-means clustering and the Hawkes process over data about three countries (Mexico, India, and Israel) to perform this study. The main result was the temporal clustering of pandemic-related demonstrations as a common feature in the studied countries.

Sengupta et al. [25] propose K-means clustering to find the similarities between the two most affected districts in India by exploring two features: population density and specialty hospitals. As a result, similar groups of the analyzed districts could be ranked based on the burden placed on the healthcare system in terms of the number of confirmed cases, population density, and the number of hospitals dedicated to specialized COVID-19 treatment.

Li et al. [14] explore the use of clustering (K-means) and classification techniques (decision tree) to find characteristics of patients infected with COVID-19. Two hundred twenty-two patients from Wuhan city were studied, and two groups were formed (common type and high-risk type).

Alanezi et al. [1] propose a Twitter sentiment analysis to determine the impact of social distance on people during the COVID-19 pandemic, making a comparison between the K-means clustering and Mini-Batch K-means clustering approaches. Two datasets (English and Arabic) were used. Results showed that based on the word frequency, the people of Italy and India were more optimistic than people from other countries during the pandemic.

Yang, Q. et al. [26] employ a decision tree model to predict death casualties in severe patients using training data with two classes (395 survivors and 57 non-survivor). Demographic, clinical, and laboratory features were used. As a result, the decision tree found that male COVID-19 patients were more prone to experience severe illness and death. Clinical characteristics and laboratory examinations were significantly different between severe and non-severe groups and between survivors and non-survivors.

From the foregoing, it can be inferred that clustering methods are widely used as a basic tool to explain differences between groups in a population or data set, including the study of the impact of COVID-19 in various fields.

## 3 Cluster Model

Clustering is an unsupervised learning technique whose goal is to find or discover groups or partitions in datasets or object collections [22]. These partitions are usually called groups so that the objects that belong to the same group are similar to each other and dissimilar to the objects of the other groups [9]. Clustering is one of the essential tasks in data mining, and analysis [24]; it has been widely used in anomaly detection and identification of outstanding features in datasets in different areas of knowledge such as biology, anthropology, materials science, medicine, statistics, and mathematics, to name a few [7, 23]. A wide variety of clustering methods have been developed since their beginnings in the 1950s [13, 15], which have been divided into two groups: partitioning and hierarchical. Despite its age, the K-means method is one of the most widely used partitioning algorithms, and today it is the de facto standard for exploring and classifying unknown datasets [11].

### 3.1 *K*-means Algorithm

Most of the partition algorithms are based on the optimization of a criterion function [22]; generally, this function is represented by $E$ (Eq. 1) for $K$-means, and its value depends on the partitions or groups ($C_i$) in the dataset $\mathbf{X}$.

$$E = \sum\nolimits_{i=1}^{K} \sum\nolimits_{x \epsilon C_i} \|x - m_i\|^2, \tag{1}$$

where $E$ is the sum of the squared error of all the objects in the dataset $\mathbf{X}$, and the centers (or means) $\boldsymbol{m}_i$ (Eq. 2) of the group $C_i$, $\mathbf{x}$ is a point in the space that represents a given object in a multidimensional space [9], and

$$\boldsymbol{m}_i = \frac{1}{\|C_i\|} \sum\nolimits_{j}^{\|C_i\|} \boldsymbol{X}_j. \tag{2}$$

The operation of the $K$-means algorithm starts by selecting or calculating $k$ centers or initial means $\boldsymbol{m}_i^0$; depending on the selection criteria, commonly $k$ objects are randomly taken from $\mathbf{X}$. Next, each object $\boldsymbol{x}_j \epsilon \mathbf{X}$ is assigned to its closest center $\boldsymbol{m}_i$. Subsequently, new centers or means $\boldsymbol{m}_i$ (Eq. 2) are calculated until the algorithm converges to a minimum value of $E$ (Eq. 1) or up to a maximum of repetitions $Q$, which are established at the beginning of the procedure, i.e., this process is repeated until $\left\|E^{(q)} - E^{(q-1)}\right\| < \Delta \; or \; q = Q$, where $q = \{1, 2, \ldots, Q\}$ and corresponds to the repetition number at that moment. This process is explained in detail in Algorithm 1.

Algoritmo 1. Pseudo-code of the $K$-means algorithm

**Input:** Dataset **X**, number of groups ($k$);
**Output:** Obtained groups $\{C_1, C_2, \ldots, C_k\}$;

1. Set the convergence criterion: Minumum error ($\Delta = 0.0001$) and máximum number of repetitions (Q = 5000);
2. Randomly assign $k$ objects of **X** as centers or initial means ($\boldsymbol{m}_i$);
3. $q = 1$;
4. **repeat**
5. $\quad$ $\boldsymbol{for}\ i = 1\ \boldsymbol{to}\ k\ \boldsymbol{do}$
6. $\quad\quad$ $\boldsymbol{x}^* \leftarrow \boldsymbol{Min}\ dist(\boldsymbol{x}_j, \boldsymbol{m}_i)$; // $\boldsymbol{x}_j \in \boldsymbol{X}$.
   //$\boldsymbol{x}^*$ is the subset of nearest neighbors $\boldsymbol{x}_j$ to $\boldsymbol{m}_i$
7. $\boldsymbol{C}_i \leftarrow \boldsymbol{x}^*$; // Assign the objects $\boldsymbol{x}^*$ to their nearest center
8. $\quad$ **end for**
9. $\quad$ $\boldsymbol{for}\ i = 1\ \boldsymbol{to}\ k\ \boldsymbol{do}$
10. $\quad\quad$ $s = ||C_i||$; // Calculate the new centers
11. $\quad\quad$ $\boldsymbol{m}_i = \frac{1}{s}\sum_j^s \boldsymbol{X}_j$ ; //$\boldsymbol{x}_j \in \boldsymbol{C}_i$ .
12. $\quad$ **end for**
13. $\quad$ q++;
14. **until** $\{(||E^{(q)} - E^{(q-1)}|| < \Delta)\ |\ |(q = Q)\}$;

### 3.2 Linkage Rules

Several methods for evaluating the dissimilarity between clusters exist, which are identified as Linkage Rules (LR) [22]. LR allow to evaluate the cluster performance. In this work, we use the distance between samples $\boldsymbol{x}_j$ of same group $C_i$ in its center $\boldsymbol{m}_i$, to evaluate the intra-group-error (Eq. 3):

$$intraE_i = \sum\nolimits_j^{\|C_i\|} dist\left(\boldsymbol{x}_j, \boldsymbol{m}_i\right); \tag{3}$$

To assess the inter-group-error, we use the distance between different centers $dist\left(\boldsymbol{m}_j, \boldsymbol{m}_i\right)$, where $i \neq j$) as it is presented in Eq. 4.

$$interE_{i,j} = dist\left(\boldsymbol{m}_i, \boldsymbol{m}_j\right) \tag{4}$$

## 4 Experimental Set up

In this section, the experimental details to allow the proper replication of the results of this research and to support the conclusions, are described.

### 4.1 Dataset

The dataset used for this work contains 1282 instances collected through questionnaires applied to students at the National Institute of Technology of Mexico, Campus Toluca. It is worth mentioning that to maintain the private data and security of each student; questionnaires were anonymous. All personal data were also omitted: date of birth, identification number, address of their permanent residence, and other private information elements.

After applying the questionnaire, the answers in a text format were transformed into a numerical form, so all data have the same data type to facilitate the work. This was done by assigning arbitrary numbers to each response; for example, if the answer is closed, 0 was given to identify a "no" and 1 for a "yes," while 2 was used for "maybe" 0 "more than 1". In the case of the sex section, 1 was set for women and 0 for men. The number label was assigned according to the total number of possible answers, starting from 0; if the question got four different answers, the labels should be 0, 1, 2, 3. At the end of applying the labels to the data, 43 attributes were obtained, which will be used to apply the K-means algorithm.

Likewise, to obtain more details about the information considered for the task and to be able to determine which elements influenced the taking of virtual classes from March 2020 to July 2021, the data collected was organized into seven categories, establishing a label for each attribute and its description as shown in Table 1.

**Table 1.** Descriptive table of the 43 attributes with their respective labels organized by category.

| Category | Attribute label | Attribute description |
|---|---|---|
| Academic | A0 | School grade |
| | A3 | Current semester |
| | A4 | Career |
| | A5 | Number of subjects currently taken |
| | A6 | Number of subjects taken last semester |
| | A7 | Number of subjects dropped in the current semester |
| | A8 | Did you suspend a semester between March 2020 and July 2021? |
| | A10 | Number of class hours dedicated per week |
| Personal | A1 | Sex |
| | A2 | Age |
| | A14 | Do you receive an external economic support to family income? |
| | A15 | Do you have a job? |
| | A16 | Civil status |
| | A17 | Number of children |
| Perception | A9 | Internet service satisfaction level |
| | A11 | Do you think preventive confinement affected in your learning quality? |
| | A12 | Do you think preventive confinement affected communication with your teachers and classmates? |

(*continued*)

**Table 1.** (*continued*)

| Category | Attribute label | Attribute description |
|---|---|---|
| | A13 | Do you think the level of knowledge acquired during the online classes was deficient? |
| Health | A18 | Do you suffer from any degenerative or chronic disease? |
| | A19 | Do you suffer from any disability? |
| | A20 | Do you suffer or have you suffered from COVID-19? |
| | A21 | Does anyone in your family have or has had COVID-19? |
| | A22 | Have you lost someone in your family due to COVID-19? |
| Tools for online classes | A23 | Computer |
| | A24 | Internet service |
| | A25 | Light |
| | A26 | Video camera |
| | A27 | Microphone |
| | A28 | Photocopier |
| | A29 | Smartphone |
| Common problems | A30 | Absence of the teacher |
| | A31 | Too much homework |
| | A32 | Lack of dynamism in the classes |
| | A33 | Lack of study spaces |
| | A34 | Little opening of subjects |
| | A35 | Technical difficulties |
| | A36 | Lack of organization of the teacher |
| | A37 | Schedule overlap between school and personal activities |
| | A38 | Acquired knowledge is deficient |
| | A39 | Lack of proper computer equipment |
| Psychological effects | A40 | Stress |
| | A41 | Anxiety |
| | A42 | Depression |

### 4.2 *K*-means Models Performance

To test the K-means model, we use the g-mean metric of intra-clusters distance (Eq. 5); it allows to have a simple measure to test the K-means performance in all clusters obtained identified by it, i.e., in the only value we can see the overall performance of the K-means algorithm.

g-mean is characterized by being sensitive to a local performance; in other words, whether some value used by this metric is low or high, they are reflected in the final value. In this work, the g-mean is computed using distances between samples of the same group, relative to its mean or centroid, as is presented in the following equation:

$$g - mean = \sqrt[k]{intraE_1 * intraE_2 * intraE_3 *, \ldots, intraE_k} \tag{5}$$

where $k$ is the total number of clusters and each intra-group-error by cluster is represented as $intraE_k$ (Eq. 3). It is worth mentioning that low values of the g-mean imply a better K-means model performance than high values.

## 5 Results and Discussion

In this section, the main results obtained in the experimentation with $K$-means algorithm are presented.

Figure 1 shows the g-mean values obtained after testing with different numbers of clusters, where big red points represent the low g-mean values for $k$ clusters. To measure the performance of each cluster, we employ Eq. 3, which asses the distance between samples of the same group and its center; then we use the g-mean metric (Eq. 5) to obtain a single error value, to determine how effective is the algorithm with $k$ different values. We choose the g-mean metric because it is sensitive to the classifier's performance on each cluster; for example, if one of the clusters has a high error, it is reflected in the g-mean value. Thus, it ensures that the lowest g-mean error corresponds with the best K-means performance on all clusters.

In accordance with Fig. 1, the clusters with the lowest errors are 6, 10, 17, 27, and 32. However, to simplify the analysis in the resultant graphics (see Figs. 2, 3 and 4 we only studied the first three clusters: 6, 10, and 17. Plotting every cluster implies very saturated background images, which could avoid focusing on the data analyses. Finally, experimental results presented in Fig. 1 exhibit that a larger number of clusters implies a decrease in the g-mean value, which may indicate that the clustering algorithm is working better.

Figures 2, 3 and 4 exhibit the behavior of the 43 attributes on each centroid (which correspond to an individual cluster), and each line in the graphics represents the value of each attribute in its corresponding centroid or cluster. To analyze these figures, we focus on attributes where the lines show different behavior; for example, in Fig. 2, attribute 7 presents very similar values in all centroids, whereas in attribute 33 exists very different values in each centroid. In other words, the comportment of this group (or cluster) of people is different. Thus, it allows the discrimination of the attributes that describe the behavior of distinct groups or clusters.

In addition, it is observed that there are lines that stand out more because it was considered to highlight the clusters with a more marked difference from others; for example, in the six test centers, the cluster “c2” and “c6” are more different from the rest, while in the ten test centers the cluster “c2”, “c7” and “c9” make the difference. Finally, the test of the seventeen centers shows that the most different clusters are “c3”,

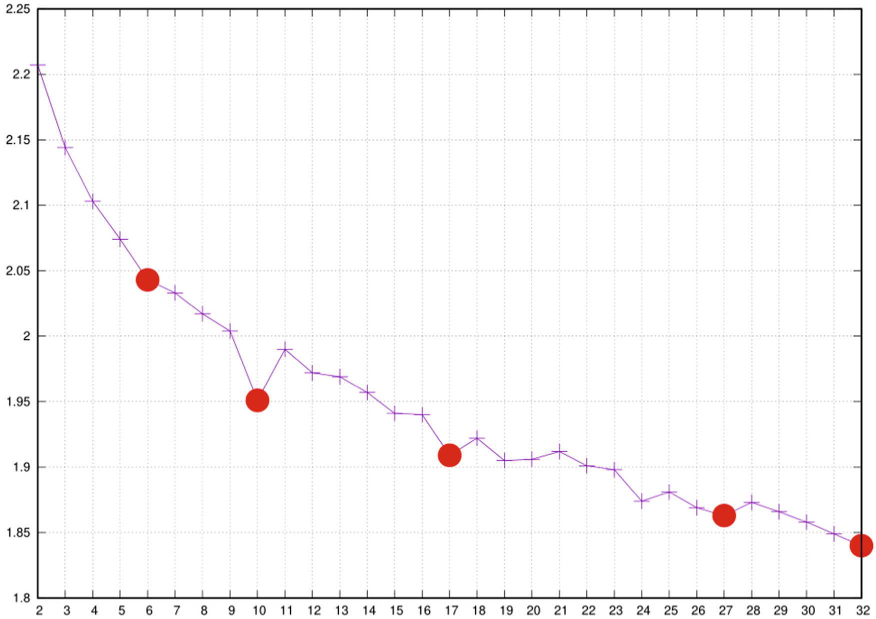


**Fig. 1.** *g*-mean (*y* axis) of *intraE* (Eq. 3) for different values of *k* (*x* axis).

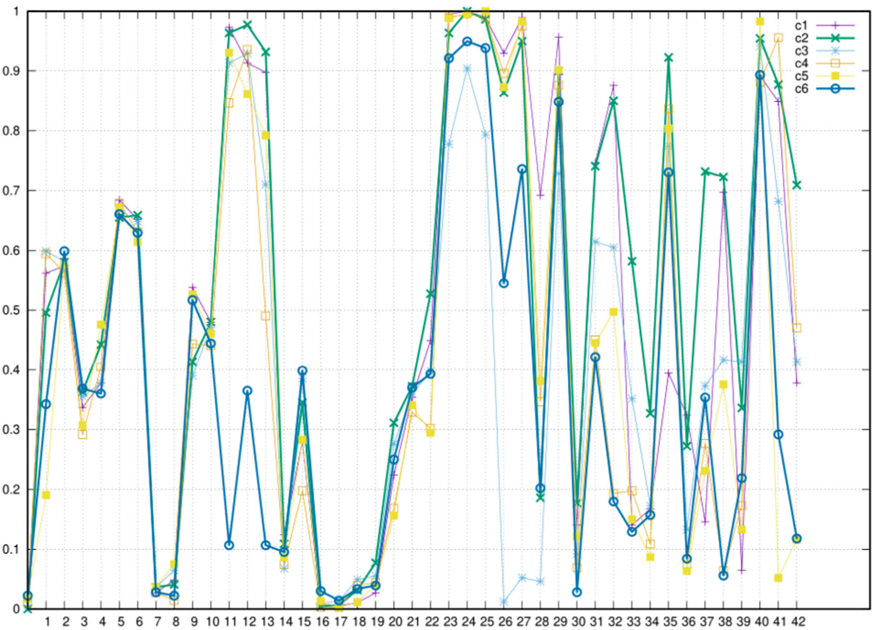


**Fig. 2.** Results obtained by the K-means model using six centers

"c7", "c13", and "c17". A similar analysis is commonly performed in others works, but using other concepts like Parallel Coordinates [21].

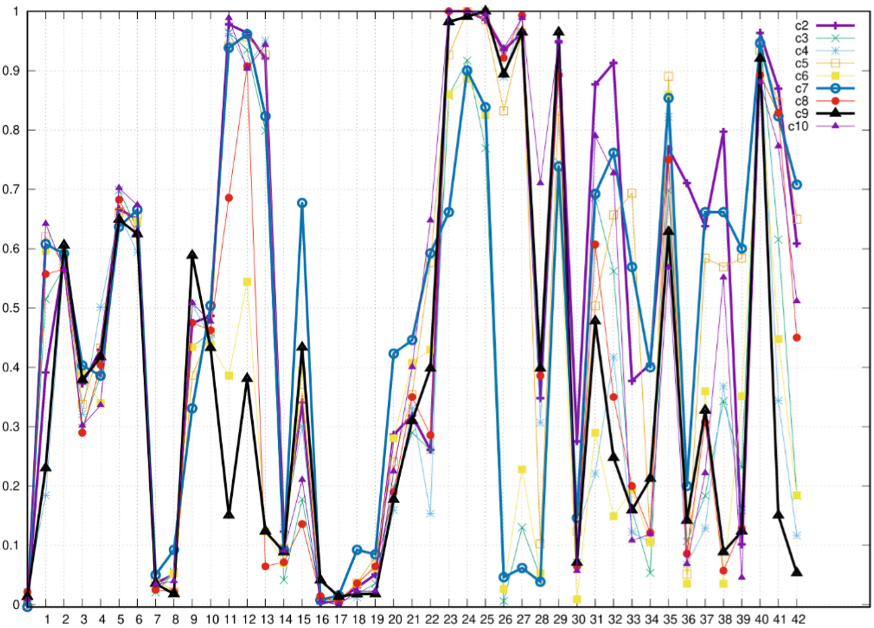


**Fig. 3.** Results obtained by the K-means model using ten centers

In Figs. 2, 3 and 4 is observed that some attributes such as A1, A11, A12, A13, A21, A22, A23, A26, A31, A32, A33, A34, A37, A38, A39, A41, and A42 are the ones that show the greatest difference between clusters. Taking into account the Table 1 and the previously highlighted attributes, the categories involved are "personal", "perception", "health", "tools for online classes", "common problems" and "psychological effects". However, due to the number of attributes included in each category, some categories stand out more than others, such as "perception", "common problems", and "psychological effects". These categories show which aspects were the most relevant for the students who took classes virtually from March 2020 to July 2021. In other words, the results show that the students were affected by the common problems that emerged while taking virtual classes, perceiving a poor quality in their learning and little communication between teachers and classmates, the same consequences that brought psychological effects such as anxiety and depression.

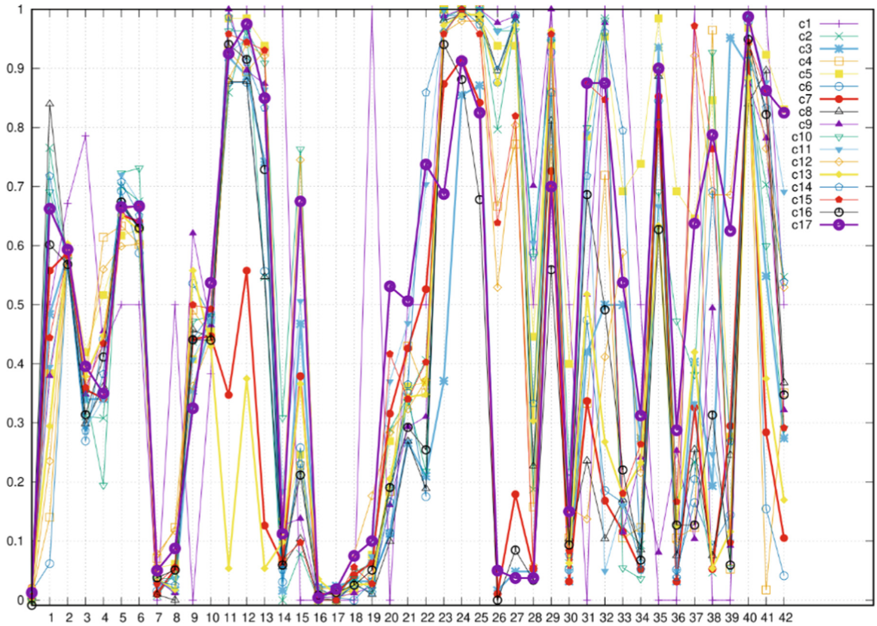


**Fig. 4.** Results obtained by the K-means model using seventeen centers

## 6 Conclusion

In this paper, we present an analysis of how preventive confinement affected the academic life of students at the National Technological Institute of Mexico, Campus Toluca, applying a clustering technique with the K-means algorithm and evaluating the performance with the g-mean metric. Experimental results expose that the test of seventeen clusters has a lower value of error using g-mean; thus, this result indicates that when a greater number of groups exist, the performance of the algorithm is better.

From a qualitative viewpoint, the results shown in the figures discussed in Sect. 5 evidenced the affectations that the students perceived during the preventive confinement: they seeming low quality of the educative system, and their level of knowledge was poor, the communication with their teachers and classmates was minimal, so reaching an agreement was a difficult task for those involved. Also, the shortage of adequate computing and communications equipment and the lack of comfortable study spaces generated widespread difficulties in performing an effective online environment. These factors could cause psychological effects in the students, such as anxiety and depression, generating more problems in their academic life.

In this work, the K-means clustering method has been used as a tool to evaluate university students' perceptions, and the quantitative results exhibit a good agreement with the qualitative insight of the students. In future work, more profound investigations could be addressed to evaluate other clustering techniques, considering the results of this study as a reference, based on the universal validity of this clustering method, which is widely recognized in the literature as the benchmark algorithm.

## References

1. ArunKum Alanezi, M.A., Hewahi, N.M.: Tweets sentiment analysis during covid-19 pandemic. In: 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), pp. 1–6 (2020). https://doi.org/10.1109/ICDABI51230.2020.9325679
2. Boldog, P., Tekeli, T., Vizi, Z., Dénes, A., Bartha, F.A., Röst, G.: Risk assessment of novel coronavirus covid-19 outbreaks outside china. J. Clin. Med. **9**(2), 571 (2020). https://doi.org/10.3390/jcm9020571
3. Campedelli, G.M., D'Orsogna, M.R.: Temporal clustering of disorder events during the COVID-19 pandemic. PLoS ONE **16**(4), e0250433 (2021). https://doi.org/10.1371/journal.pone.0250433
4. Carracedo, S., Palmero, A., Neil, M., Hasan-Granier, A., Saenz, C., Reveiz, L.: The landscape of covid-19 clinical trials in Latin America and the caribbean: assessment and challenges. Rev. Panam. Salud Publica **44**, e177 (2020). https://doi.org/10.26633/RPSP.2020.177
5. Chen, N., et al.: Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet **395**(10223), 507–513 (2020). https://doi.org/10.1016/S0140-6736(20)30211-7
6. Cheng, V.C.C., Wong, S.-C., To, K.K.W., Ho, P.L., Yuen, K.-Y.: Preparedness and proactive infection control measures against the emerging novel coronavirus in China. J. Hosp. Infect. **104**(3), 254–255 (2020). https://doi.org/10.1016/j.jhin.2020.01.010
7. Ghosal, A., Nandy, A., Das, A.K., Goswami, S., Panday, M.: A short review on different clustering techniques and their applications. In: Mandal, J.K., Bhattacharya, D. (eds.) Emerging Technology in Modelling and Graphics. AISC, vol. 937, pp. 69–83. Springer, Singapore (2020). https://doi.org/10.1007/978-981-13-7403-6_9
8. Gostin, L.O., Wiley, L.F.: Governmental public health powers during the COVID-19 pandemic: stay-at-home orders, business closures, and travel restrictions. JAMA **323**(21), 2137 (2020). https://doi.org/10.1001/jama.2020.5460
9. Han, J., Kamber, M., Pei, J.: Data Mining Concepts and Techniques, 2 edn. Morgan Kaufmann Publishers, Waltham, Mass (2006)
10. Hawkins, R.B., Charles, E.J., Mehaffey, J.H.: Socio-economic status and COVID-19–related cases and fatalities. Pub. Health **189**, 129–134 (2020). https://doi.org/10.1016/j.puhe.2020.09.016
11. Jain, A.K.: Data clustering: 50 years beyond k-means. Pattern Recogn. Lett. **31**(8), 651–666 (2010). https://doi.org/10.1016/j.patrec.2009.09.011, http://www.sciencedirect.com/science/article/pii/S0167865509002323, award winning papers from the 19th International Conference on Pattern Recognition (ICPR)
12. Jamshidi, M., et al.: Artificial intelligence and covid-19: deep learning approaches for diagnosis and treatment. IEEE Access **8**, 109581–109595 (2020). https://doi.org/10.1109/ACCESS.2020.3001973
13. Kaufman, L., Rousseeuw, P.: Finding Groups in Data: An Introduction to Cluster Analysis. Wiley Inter-Science (1990)
14. Li, Z., et al.: Efficient management strategy of covid-19 patients based on cluster analysis and clinical decision tree classification. Sci. Rep. **11**, 9626 (2021). https://doi.org/10.1038/s41598-021-89187-3
15. MacQueen, J.B.: Some methods for classification and analysis of multivariate observations. In: Cam, L.M.L., Neyman, J. (eds.) Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281–297. University of California Press (1967)
16. McAleer, M.: Prevention is better than the cure: risk management of covid-19. J. Risk Financ. Manage. **13**(3) (2020). https://doi.org/10.3390/jrfm13030046. https://www.mdpi.com/1911-8074/13/3/46

17. Melin, P., Monica, J.C., Sanchez, D., Castillo, O.: Multiple ensemble neural network models with fuzzy response aggregation for predicting covid-19 time series: the case of Mexico. Healthcare **8**(2) (2020). https://doi.org/10.3390/healthcare8020181. https://www.mdpi.com/2227-9032/8/2/181
18. NPR: Chinese authorities begin quarantine of Wuhan city as coronavirus cases multiply (2020). https://n.pr/3vAxwBA
19. World Health Organization: Who statement regarding cluster of pneumonia cases in Wuhan, China (September 2020)
20. Patel, A., Jernigan, D.B., nCoV CDC Response Team: Initial public health response and interim clinical guidance for the 2019 novel coronavirus outbreak - United States. Morb. Mortal. Wkly Rep. (MMWR) **69**(5), 140–146 (2020). https://doi.org/10.15585/mmwr.mm6905e1
21. Rendon, E., Alejo, R., Garcia Rivas, J.L.: Clustering algorithms: an application for adsorption kinetic curves. IEEE Lat. Am. Trans. **19**(3), 507–514 (2021). https://doi.org/10.1109/TLA.2021.9447701
22. de Sá, P.M.: Pattern Recognition: Concepts Methods and Applications. Springer, Heidelberg (2001). https://doi.org/10.1007/978-3-642-56651-6
23. Sánchez, M.S., Valdovinos, R.M., Trueba, A., Rendón, E., Alejo, R., López, E.: Applicability of cluster validation indexes for large data sets. In: 2013 12th Mexican International Conference on Artificial Intelligence, pp. 187–193 (2013)
24. Saxena, A., et al.: A review of clustering techniques and developments. Neurocomputing **267**, 664–681 (2017)
25. Sengupta, P., Ganguli, B., Senroy, S., Chatterjee, A.: An analysis of covid-19 clusters in India two case studies on Nizamuddin and Dharavi from the SEIQHRF model for Nizamuddin we (October 2020). https://doi.org/10.21203/rs.3.rs-68814/v1
26. Yang, Q., et al.: Clinical characteristics and a decision tree model to predict death outcome in severe covid-19 patients. BMC Infect. Dis. **21**(1), 783 (2021). https://doi.org/10.1186/s12879-021-06478-w