ISSN 1870-4069

Analysis of the Situation of University Students Facing the Change of Modality from Face-To-Face to Online Studies Using the Clustering Technique

Erika Yunuen Morales Mateos, María Arely López Garrido, Laura López Díaz, Oscar Alberto González González

Universidad Juárez Autónoma de Tabasco Cunduacán, Tabasco, Mexico

erika.morales@ujat.mx, arely.lopez@ujat.mx, {laura.lopez, oscar.gonzalez}@dais.ujat.mx

Abstract. The objective of this work is to discover groups of university students, characterized according to the change of modality from presential studies to online studies, given the current situation of health contingency caused by the COVID-19, for this purpose, sociodemographic and academic variables were considered, as well as those related to the interaction in virtual classrooms. A questionnaire was applied to a sample population of 73 students of superior level of careers related to technology, belonging to a university in the south of Mexico. The data mining technique used was clustering under the k-means algorithm, using the Weka software. As a result, three groups with different characteristics were found, among these results, one of them was identified as the group that had to contract internet service to be able to connect to their classes, another group shared their computer equipment with their relatives, another group presented an important synchronous interaction with their professors, to mention some, in which the students coincide in their preference for face-to-face studies.

Keywords: Online mode, students, university, cluster, k-means.

1 Introduction

On March 24, 2020, the federal government published in the Official Journal of the Federation the "agreement establishing the preventive measures to be implemented to mitigate and control health risks from the SARS-CoV-2 virus (COVID-19)," one of these measures being the suspension of school activities at all levels [1].

In view of this situation, the National Association of Universities and Higher Education Institutions (ANUIES) and the Higher Education Institutions (IES) agreed to continue the activities of teaching, research and dissemination of culture with the sup- port of a wide range of technological tools: virtual classrooms, remote communication platforms, repositories, libraries and digital math, among others [2], this implied a challenge for students and teachers who had to adapt from a model of face-to-face classes to virtual or online, leading to a learning that is based on the Internet

for teacher/student interaction and distribution of class materials [3], however, not all teachers had the training to implement their classes online, since this requires knowledge and learning strategies that allow students to build their knowledge.

During the time of the pandemic, Perez-Lopez et al. [4] carried out a research to know the distance education model with the aim of analyzing the incidence of the personal and familiar context of the students in the digital equity, to know the implemented teaching model and the perception and valuation that the students make of this model, they applied a questionnaire to 548 students of the University of Extremadura, the results obtained indicate that students coming from families with a low educational level have less possibilities of using digital technologies, virtual classes consisted of uploading presentations and asynchronous interaction, they value distance education negatively due to the relationship of dedication to study with the performance obtained and that teachers did not have considerations to their personal and academic circumstances.

González et al. [5] measured the academic impact of distance education implemented by COVID-19, through the virtual learning strategies used and the fluency experiences in the use of "Moodle", 525 participants participated of which 252 were Argentinean and 273 Mexican students, were surveyed through electronic links, the results obtained indicate that the use of learning strategies in the two groups is similar and there are differences in the groups regarding the fluency experience in the use of Moodle.

Şengür [6] conducted a study using data mining and automatic learning techniques to show the relationship between academic level and gender of students with anxiety and protective behaviors due to COVID-19.

In this study, an analysis was carried out on the situation of the students in relation to the change from face-to-face classes to online classes, given the health contingency that is currently present.

For that purpose, sociodemographic and academic variables were considered, as well as those related to the interaction in the virtual classroom, among them those that allowed knowing the technological resources that students use to connect to an online class, the communication tools, the type of interaction, the result of that cycle, among others.

Thus, an online questionnaire was applied to university students and data were analyzed with the descriptive technique of clustering under the K-means algorithm, which allowed the generation of groups with particular characteristics, since this algorithm makes a segmentation that allows identifying groups of elements as heterogeneous as possible and as homogeneous as possible within each group [7].

2 Proposed Materials and Methods

2.1 Description of the Data

In this work, we show results of variables related to the situation of students when they change from face-to-face to online studies. The variables are of sociodemographic,

Table 1. Applied questionnaire items

n	Item	Variable
1	Program	V1
2	Sex	V2
3	Marital status	V3
4	In the previous school year internet contract	V4
5	Internet Service Provider Experience	V5
6	Place of connection	V5
7	Adequate space to take your classes	V7
8	Cell phone for classes	V8
9	Laptop for classes	V9
10	Share equipment with family members	V10
11	You work at	V11
12	Payment of studies	V12
13	Final average of the previous cycle	V13
14	Subjects failed in the past cycle	V14
15	Media with your teachers and/or classmates	V15
16	Teachers used asynchronous presentation and interaction	V16
17	Teachers used presentation and synchronous interaction	V17
18	Teachers used presentation without interaction	V18
19	Teachers used video and asynchronous interaction	V19
20	Teachers used video and synchronous interaction	V20
21	The teachers used videos produced by him	V21
22	The teachers used videos produced by other teachers	V22
23	The teachers used blogs	V23
24	The teachers used social networks	V24
25	The teachers used technological platforms	V25
26	Teachers used online tests	V26
27	Class preference	V27

academic and online modality type, which present students of careers in technologies of a sample population of a computer science faculty, in a university of the south of Mexico.

For convenience, an online questionnaire was applied to a total of 73 students who voluntarily and anonymously participated and who were enrolled in the past school year, that is, from February to August 2020.

2.2 Online Study Experience Questionnaire

The questionnaire includes sociodemographic, academic and online questions, such as: program they are studying, sex, if they have a job, what resources they have for their studies, the way their classes were developed under this online modality, these items can be seen in Table 1.

2.3 Data Analysis

A descriptive analysis was made to know the study population, gender, curriculum and general average.

In order to know the characteristics of the population, the clustering technique was applied with the study variables (see Table 1).

With clustering, data that present similarities among them are grouped and those belonging to different groups show notable differences [9]. The k-means algorithm was used in this study due to its simplicity and frequent use. We used the data mining tool Weka [10] that includes a variety of algorithm options for data analysis, besides being a free software.

Clustering

Clustering is a descriptive task where "natural" groups of a data set are generated. The data are grouped under the principle of maximizing the similarity between the elements of a group minimizing the similarity between different groups [11]. The purpose of the is to show concentrations in the data for efficient clustering according to their homogeneity. The grouping can be done both for cases and for variables and can be of qualitative or quantitative type. The groupings are made based on the proximity or remoteness of each other, they are based on the distance [7].

The k-means algorithm is widely used thanks to its simplicity. First it is necessary to determine the number of clusters to be generated, determined by the k parameter, and k elements are selected at random, which will be the center or average of each cluster.

Then each instance is assigned to the center of the nearest cluster according to the Euclidean distance that separates it from it. For each cluster the centroid of all its instances is calculated. These centroids are taken as the new centers of their respective clusters. The process is repeated with the new centers of the clusters.

The iteration continues until the assignment of the same instances to the same clusters is repeated, since the central points have been stabilized and will remain in variables after each iteration [9].

3 Results

In Table 2, a sample of 73 students is presented, 23 women and 50 men, the educational programs considered are seven, in the population there are 20 students of Engineering in Computer Systems (ISC), 13 of Licentiate in Computer Systems (LSC), 9 students from the Administrative Computing Engineering (IIA) program, 8 students from the Information Technology (LTI) program, 5 students from the Administrative Computing (LIA) program, 4 students from the Telematics (LT) program, and 14 students from the Master's in Learning and Knowledge Technologies (MTAC) program.

Table 2. Characteristics of the study sample population.

Variables	Values	N	%
Gender	Woman	23	31.5
	Man	50	68.5
Educational program	LIA	5	6.8
	LSC	13	17.8
	LTI	8	11
	LT	4	5.5
	ISC	20	27.4
	IIA	9	12.3
	MTAC	14	19.2

The Weka software and the k-means algorithm were used to develop this analysis. The results are shown in Table 3. In cluster 1 the students of the ISC academic program were identified, in total 35, integrated mostly by men who are single. The students in this group had already contracted internet service before the contingency, with a regular service experience, they connect to their classes from their homes, in adequate space, very frequently from their cell phones and laptops and share their laptops with their brothers or sisters, they do not work and their studies are carried out with the economic support of their parents.

As for the academic aspect, the average they obtained in the previous cycle, February-August 2020 was 8.8, they did not fail subjects significantly. The media they use for their studies do so through the Teams platform.

Regarding the interaction they had with their teachers in their online classes, they have indicated that most of their teachers have used presentations with asynchronous and synchronous interaction, only some of them presented videos with synchronous or asynchronous interaction, some of the teachers used videos elaborated by them or by other teachers, blogs, social networks and other technological platforms. Most of the teachers implemented online tests. Finally, most of the students expressed a preference for face-to-face classes.

In cluster 2 there are students from ISC's academic program, in total 21, mostly by men who are single. The students of this group contracted the internet service to be able to take classes during the contingency that started last semester, with a regular service experience, they connect to their classes from their homes, in an adequate space, very frequently using their cell phones and laptops and they do not share their laptops, they do not work and their studies are done with the economic support of their parents. Academically, the average they obtained in the previous cycle was 8.47, they did not fail subjects significantly.

The media they use for their studies do so through the WhatsApp application. For interaction they had with their teachers in their online classes, this group expresses that only some of their teachers have used presentations with asynchronous and synchronous interaction, only some presented videos with synchro- nous or asynchronous interaction, and none of their teachers used videos made by them or by

Table 3. Characteristics of the obtained cluster.

Variables	Cluster 1 (35)	Cluster 2 (21)	Cluster 3 (17)
V1	ISC	ISC	MTAC
V2	Male	Male	Female
V3	Single	Single	Single
V4	No, I had already hired him a long time ago	Yes	No, I had already hired him a long time ago
V5	Regular	Regular	Good
V6	House	House	House
V7	Yes	Yes	Yes
V8	Very often	Very often	Occasionally
V9	Very often	Very often	Very often
V10	Sisters	I do not share it	I do not share it
V11	No	No	No
V12	Parent Support	Scholarship	Scholarship
V13	8.8	8.47	9.70
V14	0.45	0.61	0.
V15	Teams Platform	WhatsApp	WhatsApp
V16	Majority	Some	Majority
V17	Majority	Some	All
V18	Some	Some	Some
V19	Some	Some	Some
V20	Some	Some	All
V21	Some	None	Some
V22	Some	None	Some
V23	Some	None	None
V24	Some	None	Some
V25	Some	Most	All
V26	Most	Most	None
V27	On-site	On-site	On-site

other teachers, blogs and social networks. Most of the teachers used technological platforms and implemented online tests. This group of students indicates that they prefer face-to-face classes.

In cluster 3, students from the MATC academic program were identified, 17 in total, consisting mostly of single women.

The students of this group had already contracted the internet service before the contingency, with a good service experience, they connect to their classes from their homes, in an adequate space, occasionally from their cell phones and very frequently from their laptops and they do not share their computer equipment, they do not work and their studies are carried out with the economic support provided by their student scholarship. In the academic field, the average they obtained in the previous cycle was

9.70, they did not fail subjects. The media they use for their studies do so through the WhatsApp application. Regarding the interaction they had with their teachers in their online classes, they have indicated that most of their teachers have used presentations with asynchronous interaction and all of them in a synchronous way, only some of them presented videos with synchronous or asynchronous interaction, some of the teachers used videos elaborated by themselves or by other teachers, none of them used blogs, some social networks and all of them used technological plat- forms. No online tests were used. Finally, most of the students expressed that they preferred the face-to-face classes.

4 Conclusions

In this work, an analysis has been developed to know the current situation of university students in view of the change of modality that was presented due to the health contingency by COVID-19. For that purpose, variables related to sociodemographic and academic issues were selected, which were later analyzed with the software Weka, using the clustering technique to find groups and identify characteristics of each one of them. Thus, with this descriptive analysis three groups were identified, two of them integrated mostly by undergraduate students and one marked by the presence of graduate students. The students of cluster 2 mostly had to contract internet services to continue and the connection they have is regular.

It is notable that cluster 3, composed of master's students, had complete synchronous interaction with their professors. Among the similarities found is that they do not work and have scholarship or support from parents to carry out their studies, in terms of interactivity some of the teachers' used presentations and videos with asynchronous and synchronous interaction, in general their grades were not affected, since it is observed that the groups have averages considered as new, since they are higher than 8.0. Finally, the students in the three groups identified indicated that they prefer face-to-face classes. It is proposed to continue this study in the current semester to make a comparison with the previous semester, corresponding to this research, and to see the result of a first experience under this modality, which was not planned, and later an experience for which both students and professors were prepared.

References

1. Secretaria de Salud: Acuerdo por el que se establecen las medidas preventivas que se deberán implementar para la mitigación y control de los riesgos para la salud que implica la enfermedad por el virus SARS-CoV2. COVID-19 Diario Oficial de la Federación (2020) https://www.dof.gob.mx/nota_detalle.php?codigo=5590339&fecha=24/03/2020.
2. Asociación Nacional de Universidades e Instituciones de Educación Superior: Respuestas de las Instituciones Públicas de Educación Superior en México para enfrentar la crisis del Covid-19 (2020) anuies.mx/media/docs/avisos/pdf/200417 115709VF_ACCIONES _SES_ COVID_19_ANUIES.pdf.

3. Abreu, J.L.: Tiempos de Coronavirus: La educación en línea como respuesta a la crisis. Daena: International Journal of Good Conscience, 15(1), pp. 1–15 (2020) http://www.spentamexico.org/v15-n1/A1.15(1)1-15.pdf.
4. Pérez-López, E., Vázquez Atochero, A., Cambero Rivero, S.: Educación a distancia en tiempos de COVID-19: Análisis desde la perspectiva de los estudiantes universitarios. RIED. Revista Iberoamericana de Educación a Distancia, 24(1) (2020) doi: http://dx.doi.org/10.5944/ried.24.1.27855.
5. González, E. I., Trujillo, V., López, J.: Estrategias y experiencia de aprendizaje ante el cambio de modalidad de aprendizaje en estudiantes universitarios argentinos y mexicanos debido al distanciamiento social por Covid-19. Educateconciencia, 27(28), pp. 67–84 (2020) http://tecnocientifica.com.mx/educateconciencia/index.php/revistaeducate/article/view/257
6. Şengür, D.: Investigación de las relaciones del nivel académico y el género de los estudiantes con los comportamientos de ansiedad y protección basados en Covid-19: un enfoque de minería de datos. Revista turca de ciencia y tecnología, 15(2), pp. 93–99 (2020)
7. Perez, M.: Mineria de Datos a través de ejemplos. México: Alfaomega (2015)
8. Hernández, R., Fernández, C., Baptista, M.: Metodología de la investigación. México: McGraw-Hill Interamericana (2010)
9. García, J., Molina, J.M., Berlanga, A.: Ciencia de datos, técnicas analíticas y aprendizaje estadístico. México: Alfaomega (2019)
10. Weka: WEKA The workbench for machine learning (2020) https://www.cs.waikato.ac.nz/ml/weka/.
11. Hernández, J., Ramírez, M. J., Ferri, C.: Introducción a la Minería de Datos. Editorial Pearson Educación (2004)