UNIVERSIDAD PERUANA UNIÓN

ESCUELA DE POSGRADO

Unidad de Posgrado de Ingeniería y Arquitectura

Modelo de clasificación basado en chatbot y algoritmos no supervisados para determinar programas de intervención psicológica en estudiantes universitarios peruanos

Trabajo de investigación para obtener el Grado Académico de Maestro en Ingeniería de Sistemas con Mención en Dirección y Gestión de Tecnologías de Información

Autor:

Joyse Baldwin Huamán Labán
Roel Dante Gómez Apaza

Asesor:

Mg. Danny Lévano Rodríguez

Co-Asesor:

Dr. Miguel Ángel Valles Coral

Lima, marzo de 2022

DECLARACIÓN JURADA DE AUTORÍA DE TRABAJO DE INVESTIGACIÓN

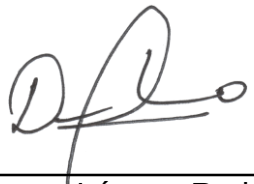
Danny Lévano Rodríguez, de la Escuela de Posgrado, Unidad de Posgrado de Ingeniería y Arquitectura, de la Universidad Peruana Unión.

DECLARO:

Que la presente investigación titulada: **"MODELO DE CLASIFICACIÓN BASADO EN CHATBOT Y ALGORITMOS NO SUPERVISADOS PARA DETERMINAR PROGRAMAS DE INTERVENCIÓN PSICOLÓGICA EN ESTUDIANTES UNIVERSITARIOS PERUANOS."** constituye la memoria que presenta el Licenciado Joyse Baldwin Huamán Labán y Licenciado Roel Dante Gómez Apaza para aspirar al Grado Académico de Maestro(a) en Ingeniería de Sistemas con Mención en Dirección y Gestión de Tecnologías de Información, cuya tesis ha sido realizada en la Universidad Peruana Unión bajo mi dirección.

Las opiniones y declaraciones en este informe son de entera responsabilidad del autor, sin comprometer a la institución.

Y estando de acuerdo, firmo la presente declaración en la ciudad de Morales, a los 23 días del mes de mayo del año 2022



Mg. Danny Lévano Rodríguez

ACTA DE SUSTENTACIÓN DE TRABAJO DE INVESTIGACIÓN

En Lima, Ñaña, Villa Unión, a 03 día(s) del mes de Marzo del año 20 22 siendo las 15:00 horas, se reunieron los miembros del jurado en la Universidad Peruana Unión Campus Lima, bajo la dirección del (de la) presidente(a): Mg. Immer Elias Cuellar Rodriguez, el (la) secretario(a): Mg. Nemias Saboya Rios, y los demás miembros: Mg. Sergio Omar Valladares Castillo y Mg. Abel Angel Sullon Macalupu y el (la) asesor(a) Mg. Danny Lévano Rodriguez con el propósito de administrar el acto académico de sustentación del trabajo de investigación titulado: "Modelo de clasificación basado en chatbot y algoritmos no supervisados para determinar programas de intervención psicológica en estudiantes universitarios peruanos" del (de la)/(los)(las) candidato(a)/s a) Joyse Baldwin Huaman Labán b) Roel Dante Gomez Apaza c) conducente a la obtención del Grado Académico de Maestro(a) en: Ingeniería de Sistemas Con mención en Dirección y Gestión de Tecnologías de Información

(Denominación del Grado Académico)

El Presidente inició el acto académico de sustentación invitando al (a la) / a (los) (las) candidato(a)/s hacer uso del tiempo determinado para su exposición. Concluida la exposición, el Presidente invitó a los demás miembros del jurado a efectuar las preguntas, y aclaraciones pertinentes, las cuales fueron absueltas por al (a la) / a (los) (las) candidato(a)/s. Luego, se produjo un receso para las deliberaciones y la emisión del dictamen del jurado.

Posteriormente, el jurado procedió a dejar constancia escrita sobre la evaluación en la presente acta, con el dictamen siguiente:

Candidato/a (a): Joyse Baldwin Huaman Labán

CALIFICACIÓN	ESCALAS			Mérito
	Vigesimal	Literal	Cualitativa	
Aprobado	16	B	Con nominación Bueno	Muy bueno

Candidato/a (b): Roel Dante Gomez Apaza

CALIFICACIÓN	ESCALAS			Mérito
	Vigesimal	Literal	Cualitativa	
Aprobado	16	B	Con nominación Bueno	Muy bueno

Candidato/a (c):

CALIFICACIÓN	ESCALAS			Mérito
	Vigesimal	Literal	Cualitativa	

(*) Ver parte posterior

Finalmente, el Presidente del jurado invitó al (a la) / a (los) (las) candidato(a)/s a ponerse de pie, para recibir la evaluación final y concluir el acto académico de sustentación procediéndose a registrar las firmas respectivas.

Presidente/a		Secretario/a
Asesor/a	Miembro	Miembro
Candidato/a (a)	Candidato/a (b)	Candidato/a (c)

ÍNDICE

Classification model based on chatbot and unsupervised algorithms to determine psychological intervention programs in Peruvian university students

Baldwin Huamán [1[0000-0002-2793-5311]], Dante Gómez [1[0000-0002-3921-7317]], Danny Lévano [2[0000-0002-1783-1105]] and Miguel Valles-Coral [3[0000-0002-8806-2892]]

[1] Escuela de posgrado, Unidad de posgrado de Ingeniería y Arquitectura, Universidad Peruana Unión, Perú
[2] Facultad de Ingeniería y Arquitectura Filial Tarapoto, EP de Ingeniería de Sistemas, Universidad Peruana Unión, Perú
[3] Facultad de Ingeniería de Sistemas e Informática, Universidad Nacional de San Martín, Perú
joyse.huaman@upeu.edu.pe

Abstract. A strategy that supports the student's academic and personal formation is that university consider tutoring as a mechanism that supports with favorable results to fight against the desertion of students. However, there are related problems in performing student segmentation and conducting psychological interventions. The objective was to formulate a classification model for intervention programs in university students based on unsupervised algorithms. For this, we carried out a non-experimental, simple descriptive study on a population of 60 university students; we carried out the data extraction process through a chatbot that applied the BarOn ICE test. After we obtained the data, the unsupervised k-means algorithm was used to group the students into sets determined based on the closest mean value obtained from the psychological test. We built a model for classifying students based on their answers to the BarOn ICE test based on K-means, with which we obtained five groups. The model classifies students by applying a different mathematical method to that used by the models applied by psychologists.

Keywords: Automated Classification, Artificial Intelligence, Grouping, K-means, University Tutoring.

1 Introduction

The emotional and psychological state of the student is a fundamental element for university academic success [1], in this sense, according to University Law No. 30220 promulgated in July 2014, in Peru the State is the regulatory agent and supervisor of compliance with the basic quality conditions in public and private universities [2]. Thus, the National Superintendence of Higher University Education (SUNEDU), as established in the Law has defined compliance with key processes to be developed to guarantee comprehensive university education [3].

Article 87 of the Law makes it mandatory for universities, through their teachers, to provide tutoring to students to guide them in their academic development and this seeks to identify, prevent and correct various situations of psycho-emotional risk, through monitoring and adequate support to the student's problem [4].

Tutoring constitutes a mechanism that can support with favorable results in the fight against the desertion of students in their university academic activities, in addition this strategy, apart from collaborating with academic training, supports personal training [5]. This process considers the university education of the student and the extracurricular needs that they may have, all developed under a personal accompaniment system, in order to support, guide and advise the student in the academic, professional and personal-social dimensions [6].

Under this premise, university tutoring together with the teaching function provide the essential mechanisms for monitoring the student's training process [7]. In this sense, universities must respond to the challenge of providing comprehensive education, institutionalizing methods and procedures that guarantee the identification of students with academic risks and establish preventive and corrective intervention programs to mitigate the probability of dropout and academic delay [8]. Thus, within universities, the tutoring area is in charge of directing certain key processes related to the psychological care and monitoring of the university student in order to promote professional development and projection [9].

However, there are factors that affect the university tutoring process such as the excessive workload of teachers, making the work they do in the academic field more difficult [10, 11]. The insufficient equipment and technological support services, where the tools used are not innovative and most of the processes are carried out manually, making it difficult to acquire new learning and adequately manage the results of psychological evaluations towards students, also affecting the time to obtain results, and limiting the performance of more periodic tests during the academic semester [12, 13].

Consequently, the difficulties that arise in monitoring the tutoring process are poor academic performance; a very common issue in universities that is related to the psychological aspect of the student [14]. Another consequence is that teachers have difficulties in virtual teaching due to technological aspects (connectivity) and this affects the basic conditions of quality, by not providing an educational service that cares about

the needs of the student [15]. That is why information and communications technology (ICT) have become a basic pillar in education today, which favors and facilitates the teaching and learning processes [16, 17]. In this sense, there is a need to incorporate ICT in universities, because it is positive due to its importance in the novelties it brings with it and the employment capacity it possesses [18, 19].

The use of ICT in the field of education has become important [20], since technological tools are linked to students, because these are the ones who use electronic equipment in their related activities in their university academic training [21]. Likewise, ICT in the university context has become an element that adapts to the needs to increase the competitive advantages of universities [22].

ICTs as computer support of the knowledge society help to build innovation capacities in the field of education [19]. Digital inclusion in organizations has allowed the development of technology based on Artificial Intelligence (AI), such as chatbots and autonomous learning systems [16, 23].

That is why AI-based technology has revolutionized the ways of conducting tutoring. This has not been excepted in university tutorials that have applied business intelligence (BI) for decision-making, however, none of these focuses on being able to identify the psychological situation when they enter university [24]. Rules-based models make it easy to create chatbots. Conversational agents or chatbots have the ability to interact with users using natural language and have the capacity for continuous learning in order to expand their knowledge in the field or subject of their programmed specialty [25].

However, it is difficult to create a bot that answers complex queries. Pattern matching is weak and therefore bots based on AIML (Artificial Intelligence Markup Language) suffer when they encounter a sentence that contains no known patterns. Also, it is quite time consuming and requires a lot of effort to write the rules manually [25]. As part of the solution to this problem, unsupervised algorithms are evaluated using various structural indices [26].

The methods used to evaluate the algorithms that solve classification problems change considerably due to the intrinsic characteristics of each type of problem [7, 27]. The main applications of unsupervised learning are related in the grouping or clustering of data [28, 29]. These algorithms are based on the distance between observations. The most used clustering algorithms are k-means clustering and hierarchical clustering [30, 31].

The development of the study considered three principles: first, the disposition of the conversational agent for the effective delivery of messages; second, to control and monitor the responses and behaviors of each student; and finally the personalized interaction with each user [32, 33]. In addition, the significant presence of issues that respond to the socio-psychological dimension is considered important [34].

This study aimed to develop an automated classification model based on the collection of information through a chatbot and the unsupervised learning algorithm application k-means, for the assertive determination of psychological intervention programs for university students carried out during the evaluation process of university tutoring of the universities of Peru.

2 Materials and methods

2.1 Unit of analysis

It is a simple descriptive, non-experimental investigation. The study used a sample of 60 students from the Faculty of Engineering and Architecture of the Universidad Peruana Unión, to whom we applied the BarOn ICE test through a chatbot, in order to obtain data on their performance regarding to the different criteria taken into account to evaluate emotional intelligence.

2.2 Chatbot construction and data extraction

We obtained data through a virtual tutoring platform based on chatbot in which a conversation was configured using the BarOn ICE test that was sent to the student's email; containing 60 items that were answered evaluating the responses on a Likert scale of five possibilities [35]. Once we obtained the data, we tabulated and processed according to the logic of the BaronICE test, generating as a result scores valued according to seven dimensions of emotional intelligence.

The chatbot is useful for automating tasks and processes in order to provide a better user experience [25, 36]. The developed chatbot can support different platforms. The prototype construction process included a review phase and an error correction phase.

The construction begins with the design of the entity-relationship model of the database that supports the logic necessary to obtain information on the dimensions of emotional intelligence of each university student. We used Navicat 15 to build the model and we deployed the database in MySQL 8.0.25.

The design of the interfaces took into account the following basic modules: importation of students, data collection instruments, scheduling of mailings, follow-up of responses and a statistical module on the performance of the use of the chatbot.

We developed the chatbot-based virtual tutoring platform in Python version 3, using the Django framework in version 3.2; it consists of the test modules and test types. We design the graphical user interface using Bootstrap 4.

2.3 Data processing

Once we consolidated and transformed the data into useful information, we take the scores of the scales obtained thanks to the BarOn ICE qualification; we carried out this process for each student, all finally stored in a dataset, for processing.

The data was compiled in the Jupyter Notebook tool using the k-means algorithm, one of the most popular clustering algorithms [37]. This algorithm assigns each data to one of the N clusters [38]. The ideal cluster in k-means is a sphere with the centroid as the center of gravity [39]. The goal in k-means is to minimize the average distance of the data from its cluster centers, where a cluster center is defined as the mean or centroid in a cluster [40, 41]. We summarized the process in Fig. 1.

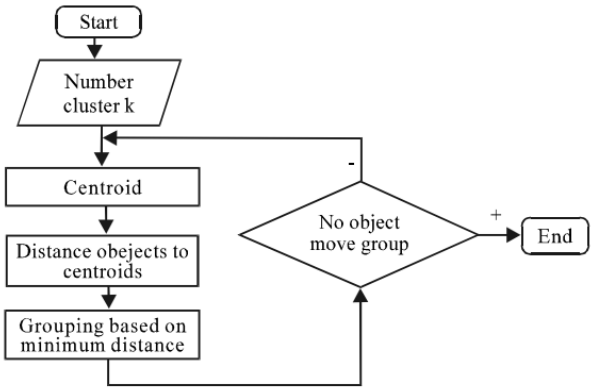


Fig. 1. K-means flow diagram [31].

The process starts with the initialization of the packages and libraries for data processing and subsequent drawing.

We used a library for mathematical calculations, for the manipulation and operationalization of numerical tables, for the generation of graphs based on lists and the one that allows the use of the k-means algorithm.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
```

To import the data, we opened the database contained in the file "BarOn Ice.csv Database" and we eliminated the first column corresponding to the identification of the students.

```
estudiantes = pd.read_csv('/home/gcode/Documentos/Base
de  datos  BarOn  Ice.csv', engine="python")
estudiantes_var = estudiantes.drop(['Alumno'],  axis=1)
```

For data scaling we import the “preprocessing” package from the “sklearn” library for the data normalization process.

```
from sklearn import preprocessing as pp
```

We assigned the normalized data to the variable "data". In addition, we copy the data from in the variable "x" for the search process for the optimal number of clusters.

```
datos_n = pp.Normalizer().fit_transform(estudiantes_var)
```

To find the optimal number of clusters, inertia is the term used to denote the sum of the squared distances of each point with respect to the centroid of its own cluster [42]. Likewise, we execute the k-means algorithm repeatedly, each time giving it a greater number of clusters in its parameters, we measure the inertia obtained in each iteration and add it to the "inertia" list.

```
inercia = []
for i in range(1, 10):
algoritmo =  KMeans(n_clusters=i,  init='k-means++',
max_iter=300,  n_init=10) algoritmo.fit(x)
inercia.append(algoritmo.inertia_)
```

Subsequently, graph the evolution of inertia with respect to the number of clusters in each iteration.

```
plt.figure(figsize=[10,6]) plt.title("Método del codo")
plt.xlabel('N°  de  Clústeres') plt.ylabel("Inercia")
plt.plot(list(range(1,  10)),  inercia,  marker='o')
plt.show()
```

Once we established the optimal number of clusters, we execute the algorithm with the initial normalized data set "data_n", and we indicated the number of clusters to generate.

```
algoritmo =  KMeans(n_clusters=5,  init='k-means++',
max_iter=300,  n_init=10)
```

Likewise, we fit the model generated with the normalized data in the variable "x" and assign the attributes "cluster_centers_" and "labels_" to the variables "centroides" and "labels" respectively.

```
algoritmo.fit(x)
KMeans(n_clusters=5)
centroides, etiquetas = algoritmo.cluster_centers_,
algoritmo.labels_
```

As a result, we print each student from the initial data set with a label that identifies the cluster to which we assigned by the algorithm.

```
print(etiquetas)
```

Later, we import the package "decomposition" from the library "sklearn" for the graphical representation of the clusters.

```
from sklearn.decomposition import PCA
modelo_pca = PCA(n_components=3); modelo_pca.fit(x)
pca = modelo_pca.transform(x)
centroides_pca = modelo_pca.transform(centroides)
colores = ['red', 'blue', 'green', 'yellow', 'black']
colores_cluster = [colores[etiquetas[i]] for i in
range(len(pca))]
plt.scatter(pca[:, 0], pca[:,1], c=colores_cluster,
marker = 'o', alpha=0.4)
plt.scatter(centroides_pca[:, 0], centroides_pca[:,1],
marker = 'x', s=100,
linewidths=3, c=colores)
xvector = modelo_pca.components_[0] * max(pca[:, 0])
yvector = modelo_pca.components_[1] * max(pca[:, 1])
columnas = estudiantes_var.columns
```

3 Results

3.1 Chatbot construction and data extraction

Fig. 2 shows the virtual tutoring chatbot interface that we built to collect data on emotional intelligence from students. We complemented the tool with different instruments provided by a psychologist; this platform carries out the application of these instruments and collects the answers given by the students. To do this, each student enters a personalized link, which we sent to his or her email address from the tutoring administration platform. As in the study of [25], the university chatbot provides efficient and accurate answers to user questions about university information. In addition, [36] develop a multiplatform chatbot that instantly answers student questions; in addition, the chatbot supports a login system to provide answers according to the different student profiles, becoming a more personalized means of communication.

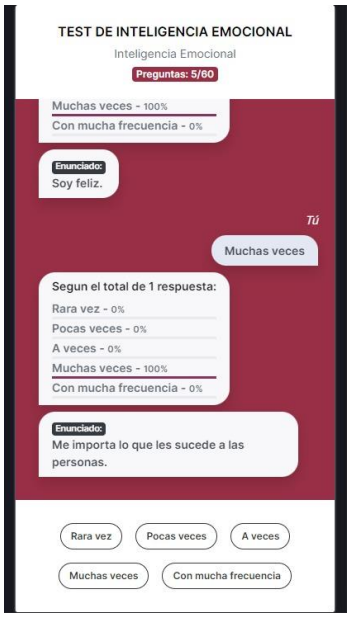


Fig. 2. Graphical interface of the chatbot on a mobile device

3.2 Data processing

As the first result of the k-means execution process in the Jupyter Notebook tool, we obtained a graph for the choice of the optimal number of clusters as shown in Fig. 3. The main idea of k-means and the algorithms Clustering is to minimize the intra-cluster variance and maximize the inter-cluster variance, where each observation must be found as close as possible to its group and as far as possible from another type of group, as stated in their studies [28, 39].

To find the optimal number of clusters, the elbow method was applied, where [33] mention that the elbow method is used to optimize the number of clusters in the k-means algorithm. Also [38] determining the exact number of clusters, in such a way the mean distance of the observations was applied to its centroid, obtaining that the value of k satisfies an increase of k.

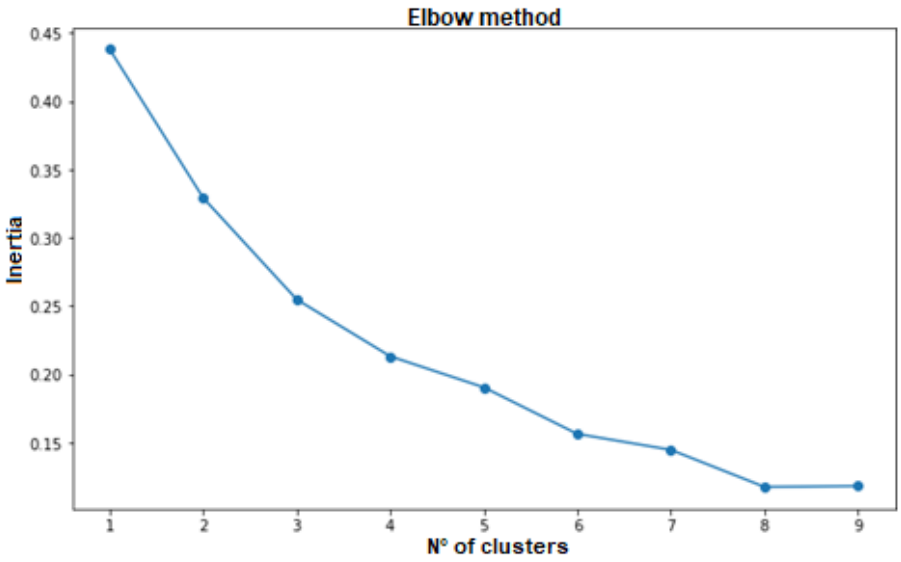


Fig. 3. Evolution of inertia with respect to the number of clusters

After the analysis of the graph of the elbow method Fig. 3, we established the optimal number of clusters and we executed the algorithm with the data set (the consolidated of the responses obtained after applying the BarOn ICE test to the group of students). Indicating the number of clusters that we have to generate and as a result, we obtained the graphical representation of the clusters as shown in Fig. 4. We have to clarify that a psychologist, who would be the commission to develop the intervention programs, as well as the effectiveness, would later analyze the resulting groupings and efficiency of the algorithm represented in the results.

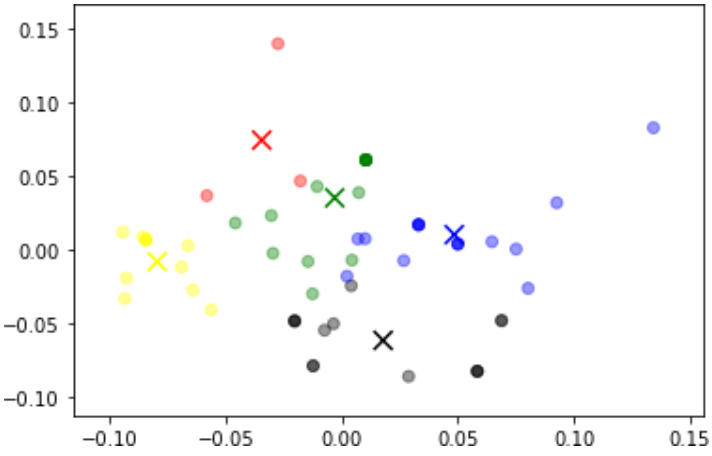


Fig. 4. Graphic representation of the clusters

Based on the results of Fig. 4, we can state that the process of segmenting the individuals of a group based on their behaviors or similar characteristics in the same segment, helps to unravel the hidden patterns in the data for a better decision making [37]. In addition [29, 37] mention that thanks to the help of the unsupervised learning algorithm k-means, the results of the clusters of the data of interest are graphically represented, helping to more optimally determine the density of the clusters, collaborating with the organization of the information generated. Considering that, the visualization tools help the researcher to achieve a better performance in the intermediate tasks of the review of the information, such as identifying main themes, associates and extracts with relevant contributions to the tasks carried out.

The application of the solution allowed generating a segmentation model of university students to determine psychological intervention programs that summarizes all the steps of the intervention carried out as shown in Fig. 5.

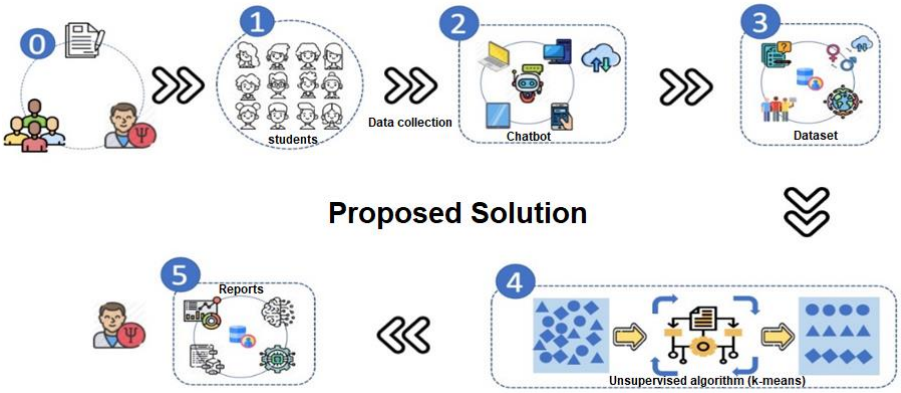


Fig. 5. Student segmentation model

In step (0), we show the entities that would participate in the intervention of the process: (1) the selection of a sample of students to whom the psychological test will be applied. (2) A chatbot was the tool built in order to apply the BarOn ICE test for data extraction. (3) Data was stored in a dataset. (4) We carried out the analysis and processing of the data, to transform them into consolidated information, where together with the support of the k-means algorithm we carried out the information segmentation process, obtaining as a result five groups with similar characteristics. (5) The intervention of a psychologist who was in charge of analyzing the final information and being able to carry out psychological intervention programs in university students. In this way, after the application of k-means groupings as in the work of [27], we can be able to identify the factors that lead to the success or failure of a student. This will allow the mentoring area to provide appropriate advice and focus more on those factors.

4 Conclusions

The classification model for university students using k-means significantly reduces the work of the psychologist in the data collection and segmentation process. Based on the processing of the responses of the applied test, clusters were obtained that must later be labeled by a qualified professional. The elbow method allows the formation of an optimal number of clusters. Due to its parameterizable characteristics, the model allows for data clustering regardless of the type of psychological test applied. It is only necessary to scale the response to obtain a number of clusters that represent logical groupings allowing covering more dimensions of the psychological field, with the purpose to identify, prevent and correct various situations of psycho-emotional risk that university students may present.

References

1. Pulido Acosta, F., Herrera Clavero, F.: La influencia de las emociones sobre el rendimiento académico. Ciencias Psicológicas. 11 (1), 29–39 (2017). https://doi.org/10.22235/CP.V11I2.1344
2. Barrios-Hernández, K. del C., Olivero-Vega, E.: Relación universidad-empresa-estado. Un análisis desde las instituciones de educación superior de Barranquilla-Colombia, para el desarrollo de su capacidad de innovación. Formación universitaria. 13 (2), 21–28 (2020). https://doi.org/10.4067/S0718-50062020000200021
3. Medina Coronado, D.: El rol de las universidades peruanas frente a la investigación y el desarrollo tecnológico. Propósitos y Representaciones. 6 (2), 703–737 (2018). https://doi.org/10.20511/PYR2018.V6N2.244
4. Chacon, M.D.: Acción tutorial en el fortalecimiento del perfil profesional universitario: aportes en el desarrollo de competencias a partir de la educación virtual. Espacios. 42 (5), 66–77 (2021). https://doi.org/10.48082/espacios-a21v42n05p05
5. Parra Silva, P.A., Ordaz Hernández, M.: Proceso de tutoría psicopedagógica. Acercamiento a la Universidad Técnica Estatal de Quevedo, Ecuador. 18 (2), 302–314 (2020). https://mendive.upr.edu.cu/index.php/MendiveUPR/article/view/1886
6. Yon Guzmán, S.E., Hernández Marín, G. del J.: Tutoría en la educación superior: análisis de la percepción de profesionales y estudiantes en una universidad pública. RIDE. Rev. Iberoam. Investig. Desarro. 9 (18), 717–747 (2019). https://doi.org/10.23913/ride.v9i18.443
7. Aguilera García, J.L.: La tutoría universitaria como práctica docente: fundamentos y métodos para el desarrollo de planes de acción tutorial en la universidad. Pro-Posições. 30 (2019). https://doi.org/10.1590/1980-6248-2017-0038
8. Sánchez Cabezas, P. del P., Luna Álvarez, H.E., López Rodríguez del Rey, M.M.: La tutoría en la educación superior y su integración en la actividad pedagógica del docente universitario. Conrado. 15 (70), 300–305 (2019). https://conrado.ucf.edu.cu/index.php/conrado/article/view/1140
9. Alonso-García, S., Rodríguez-García, A.M., Cáceres-Reche, M.P.: Analysis of the

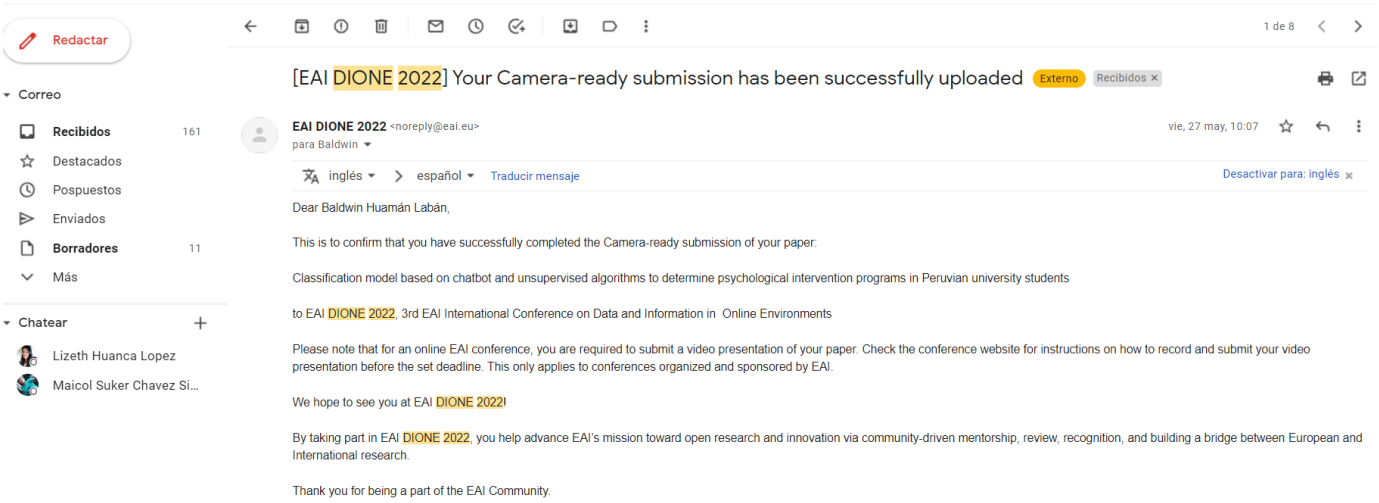
tutorial action and its impact on the overall development of the students. The case of the University of Castilla La Mancha, Spain. Formacion Universitaria. 11 (3), 63–72 (2018). https://doi.org/10.4067/S0718-50062018000300063
10. Alvites-Huamaní, C.G.: Estrés docente y factores psicosociales en docentes de Latinoamérica, Norteamérica y Europa. Propósitos y Representaciones. 7 (3), 141 (2019). https://doi.org/10.20511/pyr2019.v7n3.393
11. Mascarenhas, H., Rodrigues Dias, T.M., Dias, P.: Academic mobility of doctoral students in Brazil: An analysis based on Lattes Platform. Iberoamerican Journal of Science Measurement and Communication. 1 (3), 1–15 (2021). https://doi.org/10.47909/IJSMC.53
12. Islas Torres, C.: La implicación de las TIC en la educación: Alcances, Limitaciones y Prospectiva. RIDE. Rev. Iberoam. Investig. Desarro. 8 (15), 861–876 (2018). https://doi.org/10.23913/ride.v8i15.324
13. Poveda-Pineda, D.F., Cifuentes-Medina, J.E.: Incorporación de las tecnologías de información y comunicación (TIC) durante el proceso de aprendizaje en la educación superior. Formación universitaria. 13 (6), 95–104 (2020). https://doi.org/10.4067/S0718-50062020000600095
14. Padua Rodríguez, L.M.: Factores individuales y familiares asociados al bajo rendimiento académico en estudiantes universitarios. 24 (80), 173–195 (2019). https://www.comie.org.mx/revista/v2018/rmie/index.php/nrmie/article/view/1242
15. Montenegro Ordoñez, J.: La calidad en la docencia universitaria. Una aproximación desde la percepción de los estudiantes. Educación. 29 (56), 116–145 (2020). https://doi.org/10.18800/educacion.202001.006
16. Ocaña-Fernández, Y., Valenzuela-Fernández, L., Morillo-Flores, J.: La competencia digital en el docente universitario. Propósitos y Representaciones. 8 (1), e455 (2020). https://doi.org/10.20511/pyr2020.v8n1.455
17. Gontijo, M.C.A., Hamanaka, R.Y., Araujo, R.F. de: Research data management: a bibliometric and altmetric study based on Dimensions. Iberoamerican Journal of Science Measurement and Communication. 1 (3), 1–19 (2021). https://doi.org/10.47909/IJSMC.120
18. Vega-Hernández, M.C., Patino-Alonso, M.C., Galindo-Villardón, M.P.: Multivariate characterization of university students using the ICT for learning. Computers and Education. 121 124–130 (2018). https://doi.org/10.1016/j.compedu.2018.03.004
19. Casas-Huamanta, E.R.: Acceso a recursos tecnológicos y rendimiento académico en tiempos de pandemia y aislamiento social obligatorio. Revista científica de sistemas e informática. 2 (1), e296 (2022). https://doi.org/10.51252/RCSI.V2I1.296
20. Chen, M., Yan, Z., Meng, C., Huang, M.: The Supporting Environment Evaluation Model of ICT in Chinese University Teaching. In: Proceedings - 2018 International Symposium on Educational Technology, ISET 2018. pp. 99–103 (2018). https://doi.org/10.1109/ISET.2018.00030
21. Molinero Bárcenas, M. del C., Chávez Morales, U.: Herramientas tecnológicas en el proceso de enseñanza-aprendizaje en estudiantes de educación superior. RIDE. Rev. Iberoam. Investig. Desarro. 10 (19), e005 (2019). https://doi.org/10.23913/ride.v10i19.494

22. Gargallo Castel, A.F.: La integración de las TIC en los procesos educativos y organizativos. Educar em Revista. 34 (69), 325–339 (2018). https://doi.org/10.1590/0104-4060.57305
23. Ismail Adakawa, M.: D-Space, makerspace, and hackerspace in cyberspace: Cybersecurity strategies for digital preservation of library resources in the post-Covid-19 pandemic. Advanced Notes in Information Science. 1, 59–89 (2022). https://doi.org/10.47909/ANIS.978-9916-9760-0-5.98
24. Liang, W.: Development Trend and Thinking of Artificial Intelligence in Education. In: 2020 International Wireless Communications and Mobile Computing, IWCMC 2020. pp. 886–890 (2020). https://doi.org/10.1109/IWCMC48107.2020.9148078
25. Khin, N.N., Soe, K.M.: University Chatbot using Artificial Intelligence Markup Language. In: 2020 IEEE Conference on Computer Applications, ICCA 2020. pp. 1–5 (2020). https://doi.org/10.1109/ICCA49400.2020.9022814
26. De-La-Hoz, E.J., De-La-Hoz, E.J., Fontalvo, T.J.: Metodología de Aprendizaje Automático para la Clasificación y Predicción de Usuarios en Ambientes Virtuales de Educación. Información tecnológica. 30 (1), 247–254 (2019). https://doi.org/10.4067/S0718-07642019000100247
27. Debao, D., Yinxia, M., Min, Z.: Analysis of big data job requirements based on K-means text clustering in China. PLOS ONE. 16 (8), e0255419 (2021). https://doi.org/10.1371/JOURNAL.PONE.0255419
28. Latipa Sari, H., Suranti, D., Natalia Zulita, L., Syakur, M.A., Khotimah, B.K., S Rochman, E.M., Satoto, B.D.: Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. In: IOP Conference Series: Materials Science and Engineering. p. 012017. https://doi.org/10.1088/1757-899X/336/1/012017
29. Marisa, F., Ahmad, S.S.S., Yusof, Z.I.M., Hunaini, F., Aziz, T.M.A.: Segmentation Model of Customer Lifetime Value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM Model. International Journal of Integrated Engineering. 11 (3), 169–180 (2019). https://doi.org/10.30880/ijie.2019.11.03.018
30. Rodríguez Chávez, M.H.: Sistemas de tutoría inteligente y su aplicación en la educación superior. RIDE. Rev. Iberoam. Investig. Desarro. 11 (22), e175 (2021). https://doi.org/10.23913/RIDE.V11I22.848
31. Tinuke Omolewa, O., Taye Oladele, A., Adekanmi Adeyinka, A., Roseline Oluwaseun, O.: Prediction of Student's Academic Performance using k-Means Clustering and Multiple Linear Regressions. Journal of Engineering and Applied Sciences. 14 (22), 8254–8260 (2019). https://doi.org/10.36478/JEASCI.2019.8254.8260
32. Song, D., Oh, E.Y., Rice, M.: Interacting with a conversational agent system for educational purposes in online courses. In: Proceedings - 2017 10th International Conference on Human System Interactions, HSI 2017. pp. 78–82 (2017). https://doi.org/ 10.1109/HSI.2017.8005002
33. Marutho, D., Hendra Handaka, S., Wijaya, E., Muljono: The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News. In: Proceedings - 2018 International Seminar on Application for Technology of Information and Communication: Creative Technology for Human Life,

iSemantic 2018. pp. 533–538 (2018). https://doi.org/10.1109/ISEMANTIC.2018.8549751
34. Hernandez-Cruz, N.: Mapping the thematic evolution in Communication over the first two decades from the 21st century: A longitudinal approach. Iberoamerican Journal of Science Measurement and Communication. 1 (3), 1–10 (2021). https://doi.org/10.47909/IJSMC.88
35. Idrogo Zamora, D.I., Asenjo-Alarcón, J.A.: Relación entre inteligencia emocional y rendimiento académico en estudiantes universitarios peruanos. Revista de Investigación Psicológica. (26), 69–79 (2021). https://doi.org/10.53287/RYFS1548JS42X
36. Lee, L.K., Fung, Y.C., Pun, Y.W., Wong, K.K., Yu, M.T.Y., Wu, N.I.: Using a Multiplatform Chatbot as an Online Tutor in a University Course. In: Proceedings - 2020 International Symposium on Educational Technology, ISET 2020. pp. 53–56 (2020). https://doi.org/10.1109/ISET49818.2020.00021
37. Javed Mehedi Shamrat, F.M., Tasnim, Z., Mahmud, I., Jahan, N., Islam Nobel, N.: Application Of K-Means Clustering Algorithm To Determine The Density Of Demand Of Different Kinds Of Jobs. International Journal of Scientific & Technology Research. 9 (2), 2550–2557 (2020)
38. Liu, F., Deng, Y.: Determine the Number of Unknown Targets in Open World Based on Elbow Method. IEEE Transactions on Fuzzy Systems. 29 (5), 986–995 (2021). https://doi.org/10.1109 / TFUZZ.2020.2966182
39. Yuan, C., Yang, H.: Research on K-Value Selection Method of K-Means Clustering Algorithm. J - multidisciplinary scientific journal. 2 (2), 226–235 (2019). https://doi.org/10.3390/J2020016
40. Nainggolan, R., Perangin-Angin, R., Simarmata, E., Tarigan, A.F.: Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method. In: Journal of Physics: Conference Series. p. 012015 (2019). https://doi.org/10.1088/1742-6596/1361/1/012015
41. Br, R.W., Berahmana, S., Mohammed, A., Chairuang, K., Jimbaran, B.: Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods. Lontar Komputer : Jurnal Ilmiah Teknologi Informasi. 11 (1), 32–43 (2020). https://doi.org/10.24843/LKJITI.2020.V11.I01.P04
42. Kansal, T., Bahuguna, S., Singh, V., Choudhury, T.: Customer Segmentation using K-means Clustering. In: Proceedings of the International Conference on Computational Techniques, Electronics and Mechanical Systems, CTEMS 2018. pp. 135–139 (2018). https://doi.org/10.1109/CTEMS.2018.8769171

5. ANEXOS

Anexo 01: carta de aceptación.



Redactar

Correo

Recibidos 161
Destacados
Pospuestos
Enviados
Borradores 11
Más

Chatear

Lizeth Huanca Lopez
Maicol Suker Chavez Si...

1 de 8

[EAI DIONE 2022] Your Camera-ready submission has been successfully uploaded Externo Recibidos ×

EAI DIONE 2022 <noreply@eai.eu> vie, 27 may, 10:07
para Baldwin

inglés > español Traducir mensaje Desactivar para: inglés ×

Dear Baldwin Huamán Labán,

This is to confirm that you have successfully completed the Camera-ready submission of your paper:

Classification model based on chatbot and unsupervised algorithms to determine psychological intervention programs in Peruvian university students

to EAI DIONE 2022, 3rd EAI International Conference on Data and Information in Online Environments

Please note that for an online EAI conference, you are required to submit a video presentation of your paper. Check the conference website for instructions on how to record and submit your video presentation before the set deadline. This only applies to conferences organized and sponsored by EAI.

We hope to see you at EAI DIONE 2022!

By taking part in EAI DIONE 2022, you help advance EAI's mission toward open research and innovation via community-driven mentorship, review, recognition, and building a bridge between European and International research.

Thank you for being a part of the EAI Community.

Anexo 02: Aprobación del perfil de proyecto de tesis en formato artículo.

UNIVERSIDAD PERUANA UNIÓN
ESCUELA DE POSGRADO

"Año del Bicentenario del Perú: 200 años de Independencia"

RESOLUCIÓN N° 859-2021/UPeU-EPG-CEPG-D

Ñaña, Lima, 15 de setiembre de 2021

VISTO:

El expediente de **JOYSE BALDWIN HUAMAN LABÁN**, identificado con código Nº 200711331 y de **ROEL DANTE GOMEZ APAZA**, identificada con código N° 9610217 de la Maestría en Ingeniería de Sistemas con mención en Dirección y Gestión de TI de la Unidad de Posgrado de Ingeniería y Arquitectura;

CONSIDERANDO:

Que la Universidad Peruana Unión tiene autonomía académica, administrativa y normativa, dentro del ámbito establecido por la Ley Universitaria Nº 30220 y el Estatuto de la Universidad;

Que la Escuela de Posgrado de la Universidad Peruana Unión, mediante sus reglamentos académicos y administrativos, ha establecido las formas y procedimientos para la aprobación e inscripción del perfil del proyecto de Investigación y la designación o nombramiento del asesor para la obtención de grado de Maestro;

Que Joyse Baldwin Huaman Labán y Roel Dante Gomez Apaza, han solicitado la inscripción de su perfil de proyecto de investigación, titulado "Modelo de clasificación basado en chatbot y algoritmos no supervisados para determinar programas de intervención psicológica en estudiantes universitarios peruanos" y la designación de Asesor, encargado de orientar y asesorar la ejecución del perfil de proyecto de investigación en formato artículo;

Estando a lo acordado en la sesión del Consejo de la Escuela de Posgrado de la Universidad Peruana Unión, celebrada el 15 de setiembre de 2021 y en aplicación del Estatuto y el Reglamento General de Investigación de la Universidad;

SE RESUELVE:

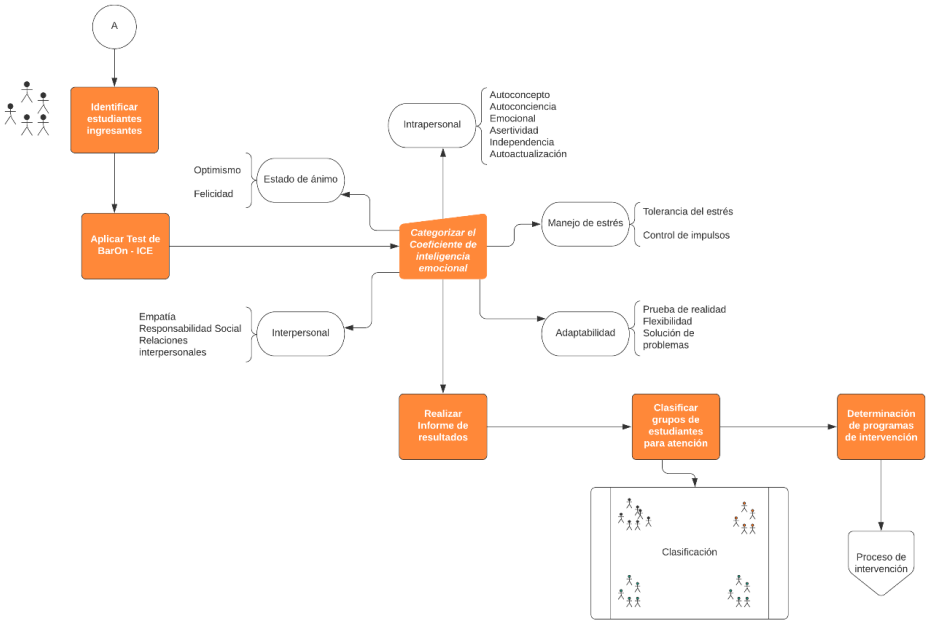
Aprobar el perfil de proyecto de investigación en formato artículo titulado "Modelo de clasificación basado en chatbot y algoritmos no supervisados para determinar programas de intervención psicológica en estudiantes universitarios peruanos" y disponer su inscripción en el registro correspondiente, nombrar al Mg. Danny Lévano Rodriguez, asesor y el Dr. Miguel Ángel Valles Coral, coasesor para que orienten y asesoren la ejecución del perfil de proyecto de Investigación en formato artículo el cual fue dictaminado por: Mg. Sergio Omar Valladares Castillo y Mg. Abel Angel Sullon Macalupu, otorgándoles un plazo máximo de doce (12) meses para la ejecución.

Regístrese, comuníquese y archívese.

Dra. Lili Albertina Fernández Molocho
DIRECTORA GENERAL

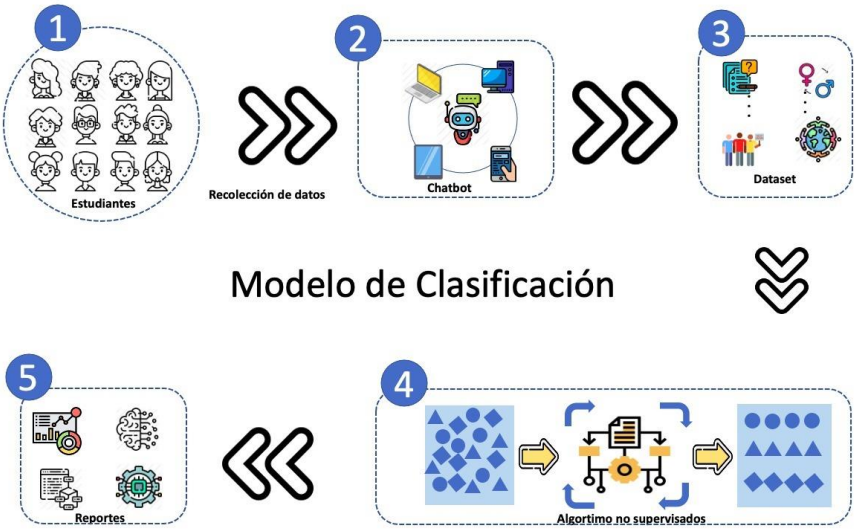
Mg. Jesús Hanco Torres
SECRETARIO ACADÉMICO

Anexo 03: Proceso que realizan las áreas de tutoría para poder evaluar, clasificar y determinar los programas de intervención a estudiantes universitarios.

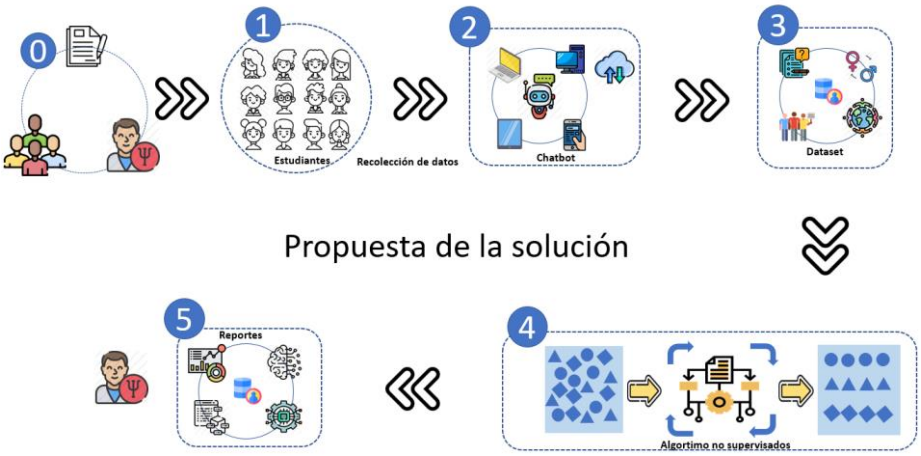


Fuente: Entrevista equipo de psicólogos UPeU, UCV, UNSM

Anexo 04:. Etapas para la implementación del modelo de clasificación.

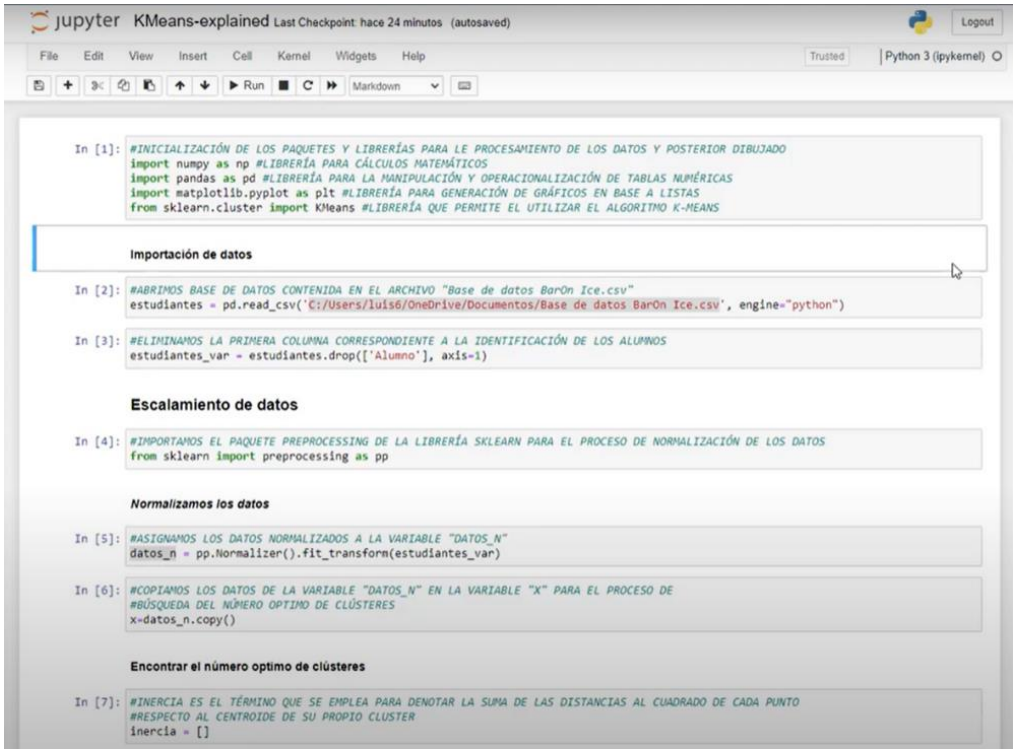


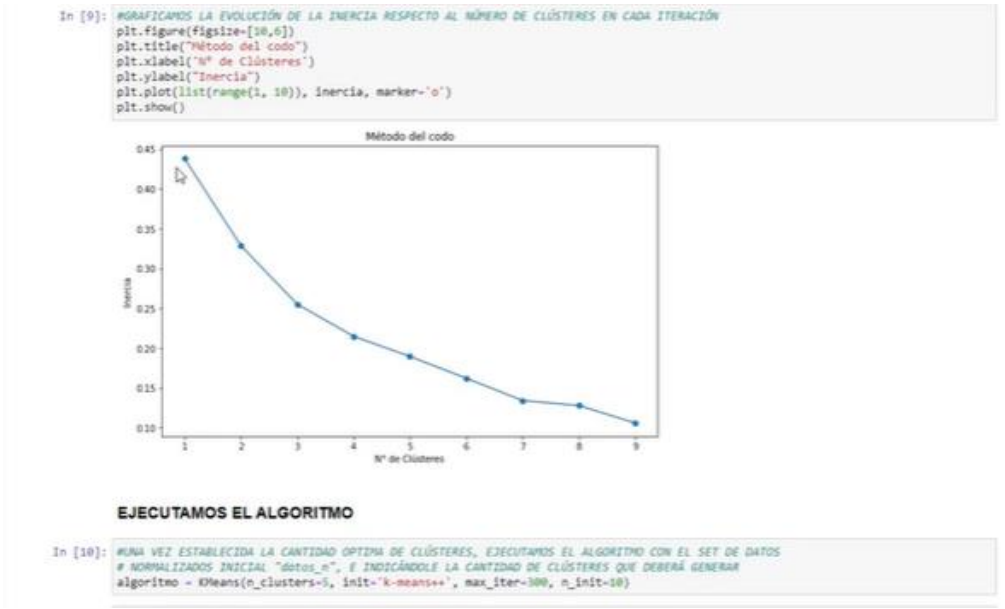
Anexo 05:. Propuesta de la solución



Anexo 06:. Interfaz gráfica del chatbot en un dispositivo móvil

Anexo 07:. Corriendo el Algoritmo





EJECUTAMOS EL ALGORITMO

In [10]:
```
#UNA VEZ ESTABLECIDA LA CANTIDAD OPTIMA DE CLÚSTERES, EJECUTAMOS EL ALGORITMO CON EL SET DE DATOS
# NORMALIZADOS INICIAL "datos_n", E INDICÁNDOLE LA CANTIDAD DE CLÚSTERES QUE DEBERÁ GENERAR
algoritmo = KMeans(n_clusters=7, init='k-means++', max_iter=300, n_init=10)
```

In [11]:
```
#AJUSTAMOS EL MODELO GENERADO CON LOS DATOS NORMALIZADOS EN LA VARIABLE "X"
algoritmo.fit(x)
```

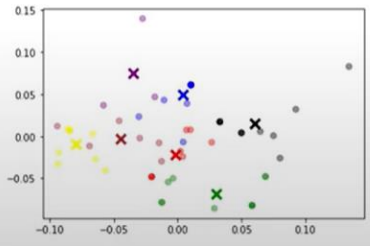
Out[11]:
```
KMeans(n_clusters=7)
```

In [12]:
```
#ASIGNAMOS LOS ATRIBUTOS "cluster_centers_" Y "labels_" A LAS VARIABLES "centroides" Y "etiquetas" RESPECTIVAMENTE
centroides, etiquetas = algoritmo.cluster_centers_, algoritmo.labels_
```

In [14]:
```
#IMPRIMIMOS CADA ALUMNO DEL SET DE DATOS INICIAL CON UNA ETIQUETA QUE IDENTIFICA EL CLÚSTER AL CUAL
# HA SIDO ASIGNADO POR EL ALGORITMO
print(centroides)
```

```
[[0.34752106 0.35778027 0.38011507 0.45470496 0.32872467 0.35899151
  0.40154247]
 [0.35128003 0.38184059 0.32885717 0.4166109  0.38134912 0.40233349
  0.37501575]
 [0.35276503 0.32104885 0.40385844 0.46889009 0.32259256 0.40079959
  0.34988642]
 [0.34621379 0.35841543 0.39383467 0.38422416 0.32960701 0.38259619
  0.4       2772459]
 [0.34962015 0.37211023 0.33300078 0.4822658  0.34830442 0.38068767
  0.3557074 ]
 [0.33560877 0.4300092  0.31455467 0.41434758 0.28659183 0.4056532
  0.42698686]
 [0.35509387 0.39623457 0.40153897 0.38399854 0.34339603 0.39641528
  0.36124056]]
```

```
pca = modelo_pca.transform(x)

#Reducción de dimensionalidad
centroides_pca = modelo_pca.transform(centroides)

#Colores para cada clúster
colores = ['red', 'blue', 'green', 'yellow', 'black', 'purple', 'brown']

#Adignación de colores
colores_cluster = [colores[etiquetas[i]] for i in range(len(pca))]

#Grafico de los componentes PCA
plt.scatter(pca[:, 0], pca[:,1], c=colores_cluster, marker = 'o', alpha=0.4)


#Grafico de los centroides
plt.scatter(centroides_pca[:, 0], centroides_pca[:,1], marker = 'x', s=100, linewidths=3, c=colores)

#Guardamos los datos en variables
xvector = modelo_pca.components_[0] * max(pca[:, 0])
yvector = modelo_pca.components_[1] * max(pca[:, 1])
columnas = estudiantes_var.columns
```



In [ ]: