International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

Predictive Multivariate Model Based on Neural Networks with Artificial Intelligence to Make Decisions in Admission Processes in Public Universities of Peru

Carlos Fidel Ponce-Sánchez[1], César Aldo Canelo-Sotelo[2]

Submitted: 19/09/2023 **Revised**: 18/11/2023 **Accepted**: 29/11/2023

Abstract: This study presents a predictive multivariate model based on neural networks with artificial intelligence for decision making in admission processes in public universities in Peru. The model has proven highly accurate in classifying applicants' grades, which highlights its effectiveness in decision-making in admissions processes. However, the importance of continuously addressing the quality and representativeness of the data used in the training of the model is highlighted to ensure unbiased results and avoid any bias or discrimination. It is suggested to carry out future research that focuses on the inclusion of diverse and representative data to achieve a more equitable and fair evaluation in the university admission processes.

Keywords: *Predictive multivariate model, Neural networks, Admission processes, Artificial intelligence, Decision making*

1. Introduction

University admission is a crucial process in the lives of students and educational institutions. Peru's public universities, as in many other countries, face the challenge of selecting the most suitable applicants and with the highest probability of academic success. In this context, having reliable prediction tools can be very useful for strategic decision making. There is a close relationship between access to high-quality education and social and economic progress, both nationally and individually. [1].

In recent years, neural networks have become a powerful tool in the field of prediction and machine learning. These networks are able to learn complex patterns and make accurate predictions from historical data sets. In particular, feedforward neural networks have been shown to be effective in predicting academic variables, such as student achievement. [2].

Likewise, an artificial neural network is a mathematical computational model that seeks to imitate the functioning of neurons in the human brain, acquiring knowledge through experiences. It consists of interconnected processing units, with the ability to identify patterns, categorize data and predict future events with great precision and accuracy. [3].

Therefore, neural network techniques, AI and machine learning are changing the field of education by improving teaching, learning and research. Neural networks can be used to analyze large volumes of educational data, such as academic records, assessments and student feedback, helping to identify patterns and trends, providing valuable information to improve teaching methods and student achievement. Importantly, the use of neural networks in a university admissions process must be carefully designed, validated and supervised to avoid bias and ensure fairness and transparency in decision-making. In addition, neural networks should be used as a complementary tool to support human decision-making, rather than completely replacing human assessment in the admissions process. [4].

On the other hand, neural networks focused on universities can be used to identify possible fraud in the admissions process. By analyzing patterns and anomalies in requests, such as false or inconsistent information, a neural network can help detect irregularities and take appropriate action.

In this way, the use of neural networks in a university's admission must be supported by sound ethics and transparency. These tools should be used fairly and equitably, taking into account the diversity and individual circumstances of applicants. They can also be used to identify possible frauds in the admission process, the use of neural networks in the educational field can be used to obtain different data and predictions that can be favorable for possible improvements in future decision making. Implementing an intelligent system based on neural networks to improve the university admissions process, using machine learning techniques and data analysis, would make more efficient and accurate decisions. [5].

*[*1]Professor , Universidad Nacional de Ingeniería, Lima-Perú*
Email ID: cponces@uni.edu.pe

[2]Professor, Universidad Nacional de Ingeniería, Lima-Perú
Email ID: ccanelo@uni.edu.pe

Nonetheless, neural networks in education have a wide range of functions, including analyzing educational data, customizing teaching, recommending learning materials, predicting student achievement, and virtual tutoring. The goal of using neural network algorithms is to improve the quality of education, adjust teaching methods to individual student needs, and create a more effective and personalized learning environment. [6].

Thus, to understand the functioning of a neural network it is essential to understand the organization of biological neurons, which are connected to each other in parallel..

However, using neural networks has advantages and disadvantages, one of the main advantages of using neural networks focused on academic performance, is accuracy in evaluation, elimination of bias, efficiency and automation, the disadvantages of using neural networks is that if the data used to train have built-in biases, it is possible for the network to reproduce and amplify them, Limited interpretability, limitations in consideration of non-academic factors, costs and resources required, implementing and maintaining a neural network infrastructure can be costly and require specialized technical resources [7].

The objective of this study is to develop a predictive model based on neural networks for the prediction of academic performance in the university admission process. This model will focus on analyzing and using applicant characteristics, such as past academic records, standardized test scores, and other relevant data.

The application of this model will allow public universities in Peru to make more informed and strategic decisions in the admission process, identifying those applicants with the highest probability of academic success. In addition, this approach can contribute to the implementation of more efficient educational policies and the adequate allocation of resources.

2. Literature Review

According to the author [8], In his research carried out at the Jorge Basadre Grohmann National University of Tacna, the use of neural networks in the admission process is proposed as a technique of automatic learning and data processing to analyze and evaluate candidates. The focus of the research is non-experimental, focusing on building a neural network that fits the input and output characteristics of the analyzed dataset. This dataset is composed of the correct answers obtained in the Admission Exam and the average grades obtained by students during the first cycle of their university career.

To carry out this process, supervised learning is used and the Re-propagation algorithm is applied in multilayer perceptron. In particular, it is observed that the intermediate layers, also known as hidden layers, perform the function of projecting the input patterns in a linearly separable manner, with the aim of minimizing errors in the generation of the output by the output unit.

According to the author [9], the use of neural networks in the educational field allows predicting the academic results of students, which provides the opportunity for the teacher to identify appropriate techniques and strategies to improve performance and competencies during the teaching and learning process. In his research, an artificial neural network (ANN) was used to predict the academic results of students in the econd of the Civil Engineering career specifically in the Playsics course at the Fabiola Salazar Leguía National Intercultural University in Bagua, Peru, using historical data.

The architecture of the ANN consists of an input layer, a hidden layer, and an output layer, and was implemented using MATLAB software. For ANN training, two algorithms available in the MATLAB Toolbox were used: the Scaled Conjugate Gradient, with a prediction rate of 70%, and the Levenberg-Marquardt, with a

prediction of 86%.

According to the author [10], to use neural networks, it is necessary to train them with data so that they can learn patterns and relationships in them. They can then be used to predict candidates' future performance and make decisions based on the established process. In his research, the author used neural networks in the validation and detection of masks in an institution. In conclusion, it is recommended to have images that present a single background in the database, as this facilitates the process of training the neural network. In addition, it is suggested to have a greater number of images to obtain more accurate results from the convolutional neural network.

According to the author [4], the use of artificial intelligence in process control is constantly increasing. Techniques such as expert systems, neural networks, genetic algorithms and fuzzy logic are used to optimize process management. These tools provide efficient solutions in situations where control using conventional techniques is difficult to implement.

In the research conducted, it was determined that the best neural network for the study was the generalized regression (GRN), according to the author [11]. Sales forecasting studies have shown that artificial neural networks (ANNs) outperform methods such as linear regression and ARIMA in accuracy. However, the author [12] points out that the skills of ANNs have not yet been fully exploited, such as their ability to associate, evaluate and recognize patterns, as well as to calculate how variables related to a company's sales can affect them.

According to the author [13], Artificial Neural Networks (ANN) are highly adaptable models that act as function approximators and have proven to be extremely beneficial in various applications related to water resources. Numerous studies have shown that neural networks outperform many of the traditional modeling methods. ANNs are a smart technology that uses experimental knowledge to solve problems in a similar way to the human brain, acquiring knowledge through a learning process and storing it in interneuronal connections, known as synaptic weights.

Despite having a limited amount of data, an ANN algorithm can produce results and, thanks to its ability to receive continuous data inputs, can be trained and remember information. One of the biggest advantages of neural networks is their ability to establish complex nonlinear models without requiring prior assumptions about the nature of the relationship. The configuration of a neural network with a hidden layer of generalized regression (GRN) is observed in the investigations.

The author [14] notes that the goal of artificial neural networks is to emulate the problem-solving process performed by the human brain within the field of artificial intelligence. These neural network models have internal variables and can be composed of multiple internal layers that establish the relationship between input and output variables. In essence, artificial neural networks mimic the behavior of the human brain, where dendrites connect to a cell nucleus that, in turn, is connected to an axon.

The author [15] indicates that the study of neural networks for the forecast of the demand of subjects is carried out in the context of universities and educational centers, where students enroll in a set of subjects, estimating the demand of students to be able to offer an appropriate number of subjects. Neural networks are a subdivision of artificial intelligence that is used in various fields of knowledge. They are frequently employed as classifiers in computer vision applications, as well as controllers and forecasting tools. The objective of these networks is to mimic the functioning of biological neurons through mathematical functions.

The author [16] in his research work on classification models to recognize dropout patterns in university students proposes to use classification models to find patterns and predict possible cases of dropout in university students An application has been developed that uses data supplied by the university and generates classification models using various algorithms, such as neural networks, ID3 and C4.5. These models use the most relevant attributes of the available information. A performance comparison was made between these models to determine which offered the best results and will be used to rank students. The results indicate that the C4.5 algorithm demonstrated significant improvements in terms of performance compared to the neural network and the ID3. The most relevant variable in the construction of the model turned out to be the relationship between the credits approved by a student and the credits he should have taken.

Likewise, the author [17] points out that the academic performance of university educational institutions, educators and students is an element of great importance in their professional training process. Educational data mining is responsible for developing models and methods to analyze data collected from educational learning environments. Through learning analysis techniques, we seek to identify patterns that allow predicting variables of interest.

Therefore, the author [18], infers that the influence of university education on entrepreneurial attitudes that The purpose of this study is to analyze how university education influences entrepreneurial attitudes through the acquisition of dynamic traits of innovation, risk propensity, personality, psychological and demographic aspects. Four dimensions composed of 75 items were used to carry out this process. Data collection was conducted on-site with students from four universities. The methodology used consisted of the use of artificial neural networks, specifically a Multilayer Perceptron (PM) model adjusted with the Economatics software. The results reveal that the development of entrepreneurship is mainly influenced by the practical training of students, which allows them to apply the knowledge acquired in companies and organizations in the region, as well as participate in business incubation within the institution. In addition, it was also observed that students earn additional income through their participation in research projects and by maintaining a discipline in sports practice.

In this way, the author [19] points out that a In an admissions process, a neural network can be used to predict students' academic performance. To accomplish this, the neural network is trained on historical data that includes relevant information about students, such as previous school grades, standardized test scores, extracurricular activities, and any other relevant data that may be available.

During training, the neural network learns to recognize patterns and relationships in this data. It uses learning algorithms to adjust the weights and connections between neurons in the network, so that it can make accurate predictions based on the input data.

The author [20], points out that a neural network for an admissions process once trained, the neural network can take the data of a student who is being evaluated in the admission process and generate a prediction of expected academic performance. This provides admissions officers with an additional tool to assess students' academic aptitude and make informed decisions about their admission to the educational institution. It should be noted that the predictions of a neural network are not foolproof and must be considered along with other factors and evaluations during the admission process.

3. Methodology

In this research, the predictive multivariate model based on Neural Networks with Artificial Intelligence is presented to make decisions in admission processes in Public Universities of Peru. The main objective of this study is to develop a predictive model that allows estimating the efficiency of candidates in university admission processes, using a variety of relevant variables.

The research focus is on applicants from the National University of Engineering, with the purpose of improving decision-making in the admission process. Through the use of Neural Networks with Artificial Intelligence, it seeks to build a multivariate model that is capable of analyzing and evaluating multiple variables related to applicants, such as academic data, exam results, extracurricular experience and other relevant factors.

The implementation of this predictive model aims to provide university institutions with an effective tool to estimate the efficiency of candidates during the admission process. This will allow more informed and objective decisions to be made, improving the quality of the selection process and increasing the likelihood of success for both students and universities.

In this article, we will detail the phases of model development, from data collection and preparation to neural network model training and evaluation. The importance of the variables used will be highlighted and the results obtained will be presented, demonstrating the usefulness and effectiveness of the proposed model.

With the predictive multivariate model based on Neural Networks with Artificial Intelligence presented in this study, it is expected to contribute to the advancement of decision-making in the admission processes in the Public Universities of Peru, providing a valuable tool to evaluate and select future students efficiently and fairly.

A. Data collection

In the data collection phase, a dataset of applicants to the admission exam of the National University of Engineering in Peru was obtained. The dataset was carefully crafted and included relevant information about applicants, such as academic background, admission test scores, and demographics.

Data were collected from 2018-2 to 2020-1, covering multiple academic periods. This allowed to have a broad and representative vision of the applicants over time. Special attention was paid to the quality of the data collected, performing validation and cleaning to eliminate outliers, missing or inconsistent data. This guaranteed the reliability of the results obtained. In addition, the inclusion of additional variables that were considered relevant for analysis and decision-making was considered. Among these variables that have been collected before resorting to the interpretation and elimination of the data are as shown in Table 1 which describes the characteristics of the data.

B. Data preprocessing

In the data preprocessing process, various techniques were applied to ensure the quality and adequacy of the data used in the study. First, a thorough cleaning of the data was carried out, removing variables or records that would not be relevant to the analysis. This allowed the study to focus on the variables of interest and ensure the consistency of the data used. In addition, a careful sample selection was made, focusing on the ordinary sample population. Through this approach, a representative group of students from the National University of Engineering was chosen, which guaranteed the validity of the conclusions obtained. The selected sample allowed inferences and generalizations about the target population, providing significant and reliable results. The steps used for a cleaning and selection of the data were as follows.

Data cleansing elimination and choice of study variables: During the rigorous process of cleaning and selecting data for our study at theNational University of Engineering, we made sure to focus on key variables that capture the essence of students´ relevant academic and socioeconomic characteristics.In this sense, we have discarded variables and records that do not provide significant information for our analysis.

In Table 2, we present the variables that have been excluded from the study, together with their respective description. These variables include personally identifiable data, details specific to the school of origin and variables related to the application process. By removing these elements, we have ensured that we protect student privacy and focus on the academic and socioeconomic aspects most relevant to our research objectives.

On the other hand, in Table 3, we show the variables that we have selected and that are being considered in our study. Each of these variables has been carefully chosen for its potential to provide valuable and meaningful information about students and their academic performance. Along with each variable, a description is included that highlights its relevance and the role it will play in our analysis.

This selective approach in the choice of variables will allow us to obtain a deeper and more accurate view of the factors that influence the admission process and the academic performance of students at the National University of Engineering. By focusing on the most relevant variables, we ensure that our findings are meaningful, reliable and applicable in the context of our research.

TABLE 1 DESCRIPTION OF THE INITIAL DATA USED IN THE STUDY

FIELD	Description		
	DATA TYPE	***Description***	***Example***
CODE	Chain	It is the students application code	20002B
paterno	Chain	It is the paternal surname of the applicant	GOMEZ
maternal	Chain	It is the maternal surname of the applicant	GARIBAY
Names	Chain	It is the first and middle name of the applicant	JOSE ARMANDO
number_Identification	Whole	It is the identity number of the student, if the student is a foreigner that number is also registered	75228234
email	Chain	The applicant's email	hell_blood_96@hotmail.com
telephone	Whole	Applicant's phone number	935825223
sex	Chain	The sex of the student	M
school	Chain	The school where the student graduated	c0439fb5b15da3415f1aa2b9f515e9df
Address_school	Chain	Address of the school which the applicant graduated	0131 MONITOR HUASCAR
ubigeo_school_dpto	Chain	Department in which the student's school is located	AVENIDA SAN MARTIN S/N
ubigeo_school_prov	Chain	Province in which the student's school is located	LIMA
ubigeo_school_dist	Chain	District in which the student's school is located	ATE
parents_school	Chain	Country of applicant's school	PERÚ
year_Exit_school	Chain	Year the applicant graduated from school	2018
speciality_of_postulacion	Chain	Professional career which you are applying for	INGENIERÍA DE TELECOMUNICACIONES
year_postulacion	Whole	Year of application	2020
cycle_postulacion	Whole	Application cycle	1
modality	Chain	Modality which postulates to the admission exam, in this case 3 types of modalities are being considered, which are: 1. Ordinary 2. National school entry 3. Extraordinary – direct income cepre	ORDINARIO
Address_candidate	Chain	Applicant's address	Calle Centauro 142

FIELD	Description		
	DATA TYPE	***Description***	***Example***
ubigeo_Address_dpto	Chain	It is the city where the applicant lives	LIMA
ubigeo_Address_prov	Chain	It is the province where the applicant lives	LIMA
ubigeo_Address_dist	Chain	It is the province where the applicant lives	LIMA
parents_birth	Chain	It is the country of birth of the applicant	PERÚ
ubigeo_birth_dpto	Chain	It is the city where the applicant was born	LIMA
ubigeo_birth_prov	Chain	It is the province where the applicant was born	LIMA
ubigeo_birth_dist	Chain	It is the district where the applicant was born	ATE
date_birth	Date	It is the date of birth of the applicant	19/10/2002
e1	Decimal	It is the first qualification with domination documented as Test 1 (P1) this test covers Academic Aptitude which includes Verbal Reasoning and Mathematical Reasoning and in Humanities which has included the subjects of: Communication, Language, Literature, History of Peru and the World, Geography and National Development, Economics, Philosophy, Logic, Psychology and Current Affairs. It comprises a minimum score of zero and a maximum of 745 points.	194.04
e2	Decimal	It is the second qualification also considered as the final qualification with documented denomination as Test2 (P2) this test covers Mathematics which have included the subjects of: Arithmetic, Algebra, Geometry and Trigonometry. Comprises a minimum score of zero and maximum of 600 points.	57.00
e3	Decimal	It is the third qualification also considered as the final qualification with documented denomination as Test3 (P3) this test covers Physics and Chemistry which have included the courses of: Physics, Chemistry, Science, Technology and Environment. Comprises a minimum score of zero and maximum of 500 points.	13.00
vocational	Decimal	It is the applicant's vocational aptitude test	0.00
Rating_final_candidate	Decimal	It is the final grade of the applicant that comes out of the calculation of the previous qualifications which will have a different interpretation of result depending on the modality. Comprises a minimum score of zero and maximum of 20.	2.862
state_acknowledged	Chain	It is the status of income of the applicant is included as a result IF in case there was income or Not in the case he had not had it.	NO

TABLE 2 DESCRIPTION OF VARIABLES ELIMINATED FROM THE STUDY

Field	**Description**
Code	The "Code" field was eliminated because it does not provide relevant information for the study, since it is an internal identifier with no direct relationship with the characteristics of the students.
Paterno	We chose to delete the "Paternal" field since paternal surname information is not considered relevant for the purposes of the study in the ordinary sample.
Materno	The inclusion of the "Maternal" field is not necessary for the analysis of the ordinary sample, so it was decided to eliminate it during the data cleansing process.
Names	The "Names" field was not used in the study, as the student name data does not provide relevant information for the purposes of the analysis.
Number_Identification	The field "Numero_identificacion" was discarded because its inclusion is not necessary for the study of the ordinary sample at the National University of Engineering.
Email	It was decided to eliminate the "Email" field since it is not relevant for the purposes of the study and does not provide significant information about the characteristics of the students..
Telephone	The field "Phone" was not considered relevant for the analysis of the ordinary sample, so it was excluded during the data cleaning process.
School	The "College" field was removed because the information about the school name is not relevant to the study during data cleansing.
Address_school	The variable was not considered due to not demonstrate a relevance to consider the direction of the school for the case of this research.
date_birth	The date of birth have brought a diversity of unique values taking a total of 2818 values within the proposed model of the neural network, therefore this variable will not be useful for this research.
vocational	Variable that has not been considered for the study because only a few specialties attended this qualification

TABLE 3. DESCRIPTION OF SELECT VARIABLES OF THE STUDY

Field	**Description**
sex	Variable that was taken into consideration because its values can give very significant results
parents_school	Variable that has been taken into consideration for the reason of being necessary for the identification of the year of graduation from the school because it can lead to making decisions due to the year of graduation.
year_Exit_school	The year of graduation from school may be relevant to understanding the educational trajectory of applicants. It can provide information about the temporality of previous studies and potentially be related to the level of academic preparation of applicants at the time of application.
speciality_of_postulacion	The application specialty can be important in analyzing applicants' preferences and areas of interest. It can help identify patterns and trends in specialization choices, as well as understand the demands and needs of the labor market in relation to areas of study.
year_postulacion	The year of application may be relevant to analyze changes or trends in admission processes over time. This can allow the identification of possible variations in admission policies, applicant demand, and other factors that may affect decision-making in admissions processes.
ubigeo_school_dpto	The ubigee of the school may be relevant to analyze geographical differences in admission results. It can help to understand if there is any variation in the results based on the geographical location of the applicants' schools.
ubigeo_school_prov	
ubigeo_school_dist	
ubigeo_address_dpto	

Field	Description
ubigeo_address_prov	The ubigee of the current address of the applicants can be useful to understand their geographical origin. It can help assess potential regional disparities in admissions results and analyze the geographic diversity of applicants. In addition, this data may be relevant to study the geographical mobility of students and its relationship with admission results.
ubigeo_address_dist	
ubigeo_birth_dpto	The ubigee of birth can provide information on the geographical distribution of applicants and assess possible differences or similarities between their place of birth and the place of study. This may be relevant for analysing student geographical mobility, regional diversity and its impact on admission outcomes.
ubigeo_birth_prov	
ubigeo_birth_dist	
parents_birth	The variable of the country of birth may be important if your study includes applicants from different countries. This would allow analysis of cultural diversity and the possible effects of nationality on admission outcomes.
e1	The grades obtained in E1, E2 and E3 represent the academic performance of the applicant in previous evaluations. These qualifications can provide information about the applicant's academic readiness and skills in specific areas. Analyzing these qualifications can help to understand the applicant's ability to succeed in their area of specialty and their level of knowledge in subjects relevant to admission.
e2	
e3	
rating_final_candidate	The applicant's final grade is a critical factor in the admission processes. This variable represents the applicant's overall performance and is critical in determining eligibility. Analyzing this variable will allow you to evaluate the relationship between final grades and admission decisions.

1) Sample selection: The selection of the sample was based on the set of applicants to the National University of Engineering. To ensure representativeness, a random sampling approach was used, where a significant number of applicants were randomly selected from each year and application cycle. In addition, the diversity of specialties was taken into account to ensure the inclusion of different areas of study. This approach allowed to obtain a sample that reflected the characteristics and distribution of the applicants in the university. By using this sample in the analysis and application of neural networks, it is expected to obtain relevant and generalizable results for the population of university applicants.
2) Treatment of missing values : By applying data imputation methods, it was possible to estimate and complete the missing values in the students' grades. These techniques were based on statistical models and advanced algorithms that allowed the missing values to be inferred accurately and reliably.The data imputation process not only ensured the integrity and consistency of the data, but also made it possible to make the most of the information available in the dataset. By filling in the missing values, the structure and relationships between the variables could be preserved, providing a more complete and accurate view of students' grades.
3) Normalization of numerical variables : In the stage of normalization of numerical variables, techniques were applied to standardize and scale the relevant variables of the dataset.

For the variables "e1", "e2" and "e3", which are numerical, a normalization approach was used to fit the values within a common range, such as between 0 and 1. This allowed to eliminate any bias in the data and ensure that the variables contribute equitably to the analysis and subsequent modeling.

In the case of the variable "calificacion_final_postulante", also of numerical type, normalization techniques were applied to ensure that the values are on a comparable scale. This facilitated the comparison and interpretation of applicants' scores, without one variable having a disproportionate impact on the analysis due to its original scale.0

As for the variable "sex", which is of chain type, numerical normalization was not required, since it is a categorical variable that represents a qualitative characteristic. However, proper coding may have been done to represent categorical values numerically, which is common in processing data for further analysis.

Finally, the variable "ubigeo_direccion_dpto", also of string type, could have been converted into a categorical variable, assigning a numerical value to each department. This would allow their inclusion in subsequent analysis and modeling, either by coding techniques or by using dummy variables to represent each category.

The normalization of these numerical variables ensured that all variables were on a consistent and comparable scale, which facilitated the analysis and interpretation of the results obtained through the use of neural networks and other mathematical and statistical techniques.

The numerical variables were normalized to ensure that they had a similar scale, using min-max standardization or normalization techniques.

4) Analysis and detection of outliers : In the stage of analysis and detection of outliers, statistical and graphical techniques were applied to identify and treat those values that deviated significantly from the expected pattern in the dataset. This allowed us to examine and better understand the distribution of variables and assess the presence of outliers. The main objective was to identify possible errors in the measurement, incorrect data entry or rare events in the studied phenomenon that could affect the results of the analysis. When outliers were detected, their relevance was assessed and a decision was made on how to treat them based on the potential impact on data interpretation. In this case the formula for Grubbs' statistical calculation [21] was used as shown in equation (1).

$$G = (|X - \bar{X}| / S) / \sqrt{(n - 1) / n} \quad (1)$$

Where:

- G is the Grubbs statistician.
- X is the value suspected of being an outlier.
- $\bar{X}$ is the average of the data.
- S is the standard deviation of the data.
- n is the sample size.

The Grubbs statistic is compared to a critical value of the t-Student distribution to determine whether the X value is a significant outlier. If the calculated G-statistic is greater than the corresponding critical value, the null hypothesis that X is a typical value is rejected and considered as an outlier. This formula allows for a more rigorous and accurate analysis in the detection of outliers, taking into account both the mean and standard deviation of the data and the sample size [22].

C. Exploratory data analysis

Within the exploratory data analysis, the k-means algorithm was applied in order to determine segmentations in the e1, e2, e3 and calificación_final grades of the applicant. This technique, widely used in data exploration, allows us to identify clustering patterns that will help us better understand student performance in terms of grades.

Within the exploratory data analysis, the k-means algorithm was applied in order to determine segmentations in the e1, e2, e3 and calificación_final grades of the applicant. This technique, widely used in data exploration, allows us to identify clustering patterns that will help us better understand student performance in terms of grades.

Within the exploratory data analysis, the k-means algorithm was applied with the aim of identifying clustering patterns in the dataset. This technique, widely used in data exploration, allows for a deeper understanding of the underlying structure of data. The k-means algorithm was selected because of its efficiency in identifying clusters and its ease of implementation [23]. Through its application, we sought to group the data into k clusters, k being a previously determined value based on the context of the study.

The mathematical model used in the k-means algorithm is based on minimizing the sum of the squared distances between data points and centroids of clusters. This approach is based on Euclidean distance, a measure of similarity between points in multidimensional space.

Considering a dataset $X = \{x_1, x_2, ..., xn\}$, where each xi represents a data point in a space of dimension d, and a positive integer k indicating the number of clusters to be formed, the goal is to find k centroids $C = \{c_1, c_2, ..., ck\}$ that minimize the sum of the squared Euclidean distances between the data points and the nearest centroids.

The model can be expressed as a cost function to be minimized, called J(X, C), which represents the sum of the squared Euclidean distances between the data points and the centroids. To apply the k-means algorithm, an iterative process consisting of two steps is followed. First, each data point is assigned to the nearest centroid based on the Euclidean distance. Then, the centroids are updated by calculating the average of the points assigned to each cluster. These two steps are repeated until a convergence criterion is met, such as stabilization of centroids or maximum number of iterations.

The classification of grades is a clear example of the application of grouping techniques, which are addressed in this study. Figure 1 shows an example of the desired results of the research, which will be discussed in detail in the results section.

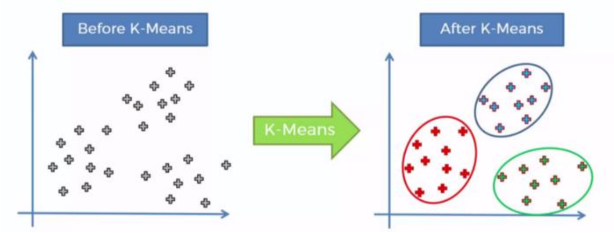

Fig. 1. Feedforward neural network model

D. Design and training of the neural network

In the stage of Design and training of the neural network, a predictive model based on feedforward neural networks was developed with the aim of predicting the level of qualification of students at the National University of Engineering. Figure 2 shows the interpretation of this network, which was used to forecast the demand for subjects.

To achieve this, a suitable neural network architecture was implemented and training was carried out using a set of historical data.

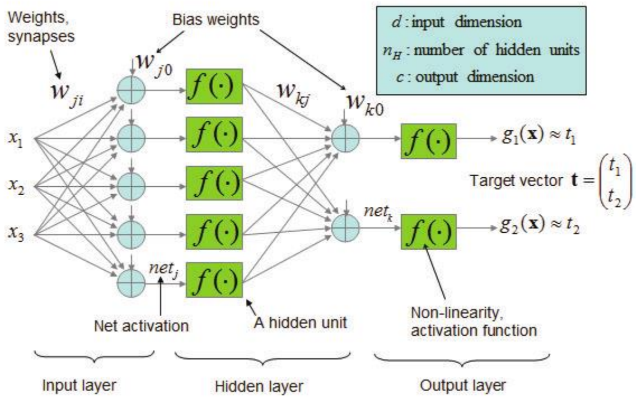

Fig. 2. Feedforward neural network model

The mathematical model is represented as follows [24]: it is considered a set of inputs $X = [x_1, x_2, ..., xn]$, where each xi represents an entry in a space of dimension d. In addition, we have a set of weights $W = [w_1, w_2, ..., wm]$, where each wj represents the weights associated with the connections of the layer j, and a set of biases $b = [b_1, b_2, ..., bm]$, where each bj represents the biases associated with the neurons of the layer j. The feedforward process is performed at each layer of the network as follows:

- Input layer: $Z_1 = XW_1 + b_1$
- Hidden layers (if any): $Zj = a(Zj_{-1}Wj + bj)$ for j = 2, 3, ..., L
- Output layer: $Zl = a(Zl_{-1}Wl + bl)$

Where a(·) represents the activation function applied to each neuron of the corresponding layer.

Validation and evaluation of the neural network

In the Neural Network Validation and Evaluation section, several steps were carried out to measure the performance and effectiveness of the neural network-based predictive model.

Initially, the mean square error (MSE) was used as a cost function to assess the discrepancy between the model predictions and the actual values of the test data [25]. The mathematical formula of the MSE is defined as:

$MSE = (1/n) * \Sigma_i (y_i - \hat{y}_i)^2$ (1)

Where n represents the total number of samples in the dataset, yi is the actual value of sample i, and ŷi is the value predicted by the model for sample i.

In addition to the SSM, the coefficient of determination (R^2) was used as an additional metric to assess the quality of the predictions [26]. The coefficient of determination measures the proportion of variability in the output data that can be explained by the model. A value of R^2 close to 1 indicates a good fit of the model to the data.

Where n represents the total number of samples in the dataset, yi is the actual value of sample i, and ŷi is the value predicted by the model for sample i.

In addition to the SSM, the coefficient of determination (R^2) was used as an additional metric to assess the quality of the predictions [26]. The coefficient of determination measures the proportion of variability in the output data that can be explained by the model. A value of R^2 close to 1 indicates a good fit of the model to the data.

Assessment metrics, such as MSE and R^2, were calculated for this test dataset.

4. Results

In this research, an analysis is performed using the clustering approach, which is fundamental for the next step in the structure of the neural network model. The main objective is to accurately segment the independent variable of the study, allowing an adequate classification of the data.

First, a thorough analysis is carried out using clustering techniques that allow us to identify patterns and group the data into homogeneous sets. This is crucial for understanding the underlying relationships and characteristics in the data.

In addition, random sampling is performed to select the most relevant and significant data in the analysis. This ensures that the results obtained are representative and reliable.

By approaching the analysis from this refined and technical approach to clustering, we achieve accurate segmentation of the independent variable, which in turn allows us to gain a deeper understanding of the data. This provides a solid foundation for informed decision-making and improving the quality of research results.

A. Segmentation Analysis and Prediction of Academic performance

Figure 3 presents the clustering of students' final grades, while Table 3 shows the grade ranges for each group. These two elements provide valuable information about the distribution of students at different levels of academic performance.

Looking at Figure 3, it can be seen that approximately 12.04% of the students sampled are in "Group 1: High Grades". This represents a total of 239 students in the sample, suggesting that a considerable proportion of students have achieved high grades, reflecting strong academic performance.

On the other hand, "Group 2: Average Grades" includes students whose grades range from 6,190 to 10,163. This group comprises about 46.57% of the students sampled, corresponding to a total of 931 students. Here, we find an intermediate level of performance, with students who have performed reasonably but do not reach the same level as those in the high-grade group.

Finally, "Group 3: Low Grades" is made up of students whose grades range from 0.00 to 6.178. This group represents approximately 41.39% of the students sampled, which equates to 827 students. Here we find a low level of performance, with students who have obtained lower grades compared to the other groups.

Academic performance group segmentation, depicted in Figure 3 and Table 4, provides a valuable predictor variable for models such as neural networks. These results allow us to identify patterns and trends that can improve the accuracy in predicting students' academic performance.

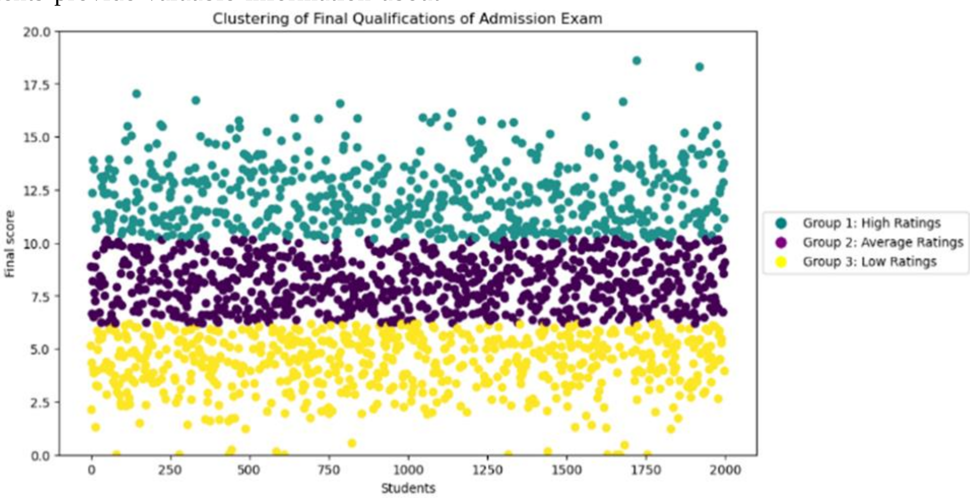

Fig. 3. Clustering of Final Qualitications of admission Exam.

TABLE I. SEGMENTATION OF QUALIFICATION OF LOW MEDIUM AND HIGH OF THE FINAL QUALIFICATION

Group name	Rango Mínimo	Rango Máximo
Group 1: High Ratings	10.178	18.592
Group 2: Average Ratings	6.190	10.163
Group 3: Low Ratings	0.00	6.178

B. According to admission test scores

Figure 4 shows the distribution of scores in variables e1, e2 and e3 for each of the identified groups. We can observe that "Group 1: High Ratings" has the highest values in all three variables, with ranges ranging from 225.80 to 636.04 for e1, 39.64 to 488.24 for e2, and 137.52 to 550.80 for e3. This indicates that students in this group, who make up about 30% of the total sample, have scored significantly higher compared to the other groups.

On the other hand, "Group 2: Average Ratings" shows intermediate values in all three variables, with ranges ranging from 198.00 to 567.00 for e1, 1 to 264.00 for e2, and 24.00 to 474.00 for e3. This suggests that students in this group, who make up about 40% of the total sample, have had moderate academic performance, not reaching the highest levels, but also not getting low grades.

Finally, "Group 3: Low Ratings" presents the lowest values in all three variables, with ranges ranging from 95.00 to 488.24 for e1, 1 to 258.00 for e2, and 13.00 to 398.00 for e3. This indicates that students in this group, who make up about 30% of the total sample, have had lower academic performance compared to the other groups.

These findings are supported by Table 5, which shows the minimum and maximum ranges of ratings for each group in variables e1, e2, and e3. These ranges provide a quantitative measure of the variability of scores in each group and support the visual observation in Figure 4. In addition, considering that the sample used represents approximately 12% of the total population of 16,607 students, these results can be generalized to the study population with a statistically significant level of confidence.

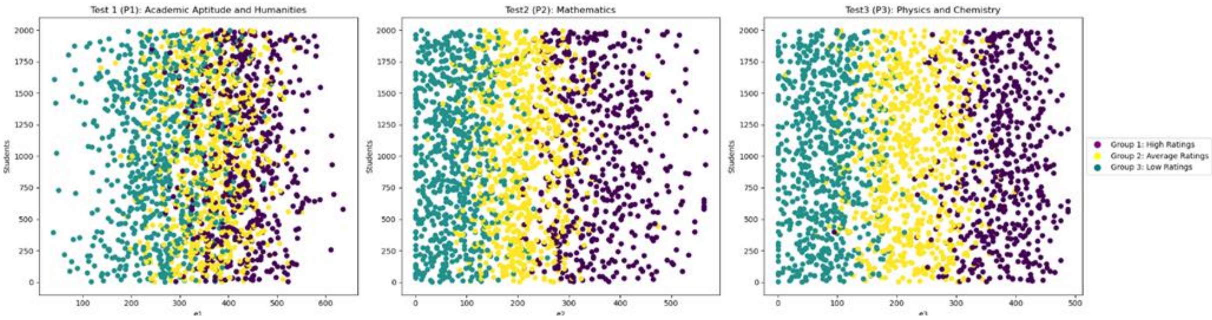

Fig. 4. Clustering e1, and e2 and e3 Qualitications of admission Exam

TABLE V. RATING SEGMENTATION OF LOW MEDIUM AND HIGH RATING E1,E2 Y E3

Exam	Group name	Rango Mínimo	Rango Máximo
e1	Group 1: High Ratings	225.80	636.04
	Group 2: Average Ratings	198.00	567.00
	Group 3: Low Ratings	95.00	488.24
e2	Group 1: High Ratings	39.64	488.24
	Group 2: Average Ratings	1	264.00
	Group 3: Low Ratings	1	258.00
e3	Group 1: High Ratings	137.52	550.80
	Group 2: Average Ratings	24.00	474.00
	Group 3: Low Ratings	13.00	398.00

Regarding the Validation and Evaluation of the Neural Network model, its performance in the admission processes of Public Universities of Peru was evaluated. The results obtained were as follows.

Coefficient of determination (R-squared): A value of 53.2% was obtained, indicating that the model explains approximately 53.2% of the variability in the final grades of the applicants.

Confounding matrix: The model achieved an accuracy of 84.4% in correctly ranking grades, demonstrating its ability.

C. To make accurate predictions and make appropriate decisions in the admissions process.

F1 score: An F1 score of 64.3% was obtained, indicating a balance between accuracy and recall in rating rankings.

The results obtained in the validation and evaluation of the Neural Network model are promising and support its usefulness in the admission processes of Public Universities of Peru. The coefficient of determination (R-squared) of 53.2% indicates that the model can explain a significant part of the variability in applicants' final grades. In addition, the high accuracy of 84.4% in correctly ranking ratings demonstrates the model's ability to make sound decisions. The F1 score of 64.3% reinforces the balance between accuracy and recall. These results will be presented in Table 6, providing a clear vision and supporting the relevance and applicability of the model in the field of university admissions.

Figure 5 shows the confusion matrix obtained during the evaluation of the Neural Network model. This matrix clearly and concisely visualizes the model's performance in ranking applicants' grades into specific categories, such as low, medium, and high. The values in the matrix represent the number of correct and incorrect predictions in each category. This graphical representation allows us to analyze and understand the performance of the model in decision-making in the admission processes. Figure 5 provides a valuable tool for evaluating and improving the accuracy and effectiveness of the model in future research.

TABLE VI. RESULT OF ACURRACY BY DIFFETENT MEASUREMENT MODELS

Group name	**Accuracy**
Coefficient of determination (R-squared):	53.2%
Confusion matrix	84.4%
Punctuation F1	64.3%

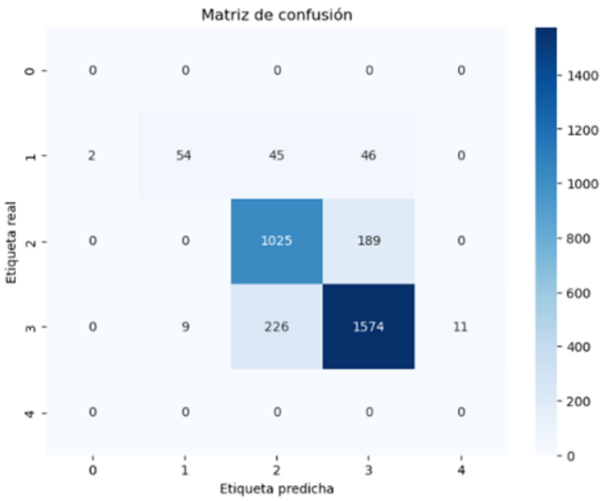

Fig. 5. Confusion matrix

Fig. 6. A simulation of predictive results will be performed using validation data as a starting point. To do this, a random sample of data will be applied as input to generate predictions using the trained Neural Network model.

Fig. 7. The random sample will be randomly selected from the available validation data, with the aim of simulating different scenarios and evaluating the performance of the model in different cases. This technique will allow us to obtain a broader and more robust perspective of predictions and their ability to adapt to different situations.

Fig. 8. Once the predictions are obtained from the random sample, a detailed analysis of the results will be performed. The distribution of predicted ratings will be examined, relevant patterns or trends identified, and compared with actual validation data.

Fig. 9. This simulation of predictive results will provide us with additional insight into the ability of the Neural

Network model to make accurate predictions and make decisions in the admission processes. In addition, it will allow us to evaluate the robustness and generalization of the model in different scenarios and validate its applicability in practical situations.

Fig. 10. Table 7 shows the results of simulations based on the validation data. These results represent the estimated academic performance for each department and are divided into grading categories: high, medium and low. In addition, additional information is provided on the gender and average rating for men in each department.

Fig. 11. It is important to note that the information presented in the table is organized by department. Each department includes information on the provinces and districts that belong to it. This allows a more detailed view of the results and their distribution within each department.

TABLE VII PREDICTION OF QUALIFICATION BY DEPARTMENT OF CURRENT EXPERIENCE AND QUALIFICATION BY MEN AND WOMEN

Department of experience	**Men Rating**	**Women Qualification**
Amazon	Media	Loud
Áncash	Casualty	Media
Apurímac	Casualty	Casualty
Arequipa	Media	Media
Ayacucho	Media	Casualty
Cajamarca	Casualty	Media
Callao	Loud	Loud
Cusco	Loud	Media
Huancavelica	Casualty	Casualty
Huánuco	Media	Casualty
Ica	Media	Media
Junín	Media	Casualty
La Libertad	Media	Media
Lambayeque	Media	Casualty
Lima	Loud	Loud
Loreto	Media	Media
Madre de Dios	Loud	Loud
Moquegua	Loud	Media
Pasco	Media	Casualty
Piura	Casualty	Media
Puno	Casualty	Media
San Martín	Loud	Media
Tacna	Loud	Loud
Tumbes	Casualty	Casualty
Ucayali	Casualty	Media

Table 8 shows the qualification of men and women in different school departments. Rating values are represented as "Low", "Medium" and "High". These values are used as an example to illustrate how ratings might be classified based on these categories.

The table provides information on the qualification of men and women in each college department. For example, in the department of Arequipa, both men and women have a rating considered "High". In the department of Apurímac, both men and women have a rating considered as "Low".

TABLE VIII PREDICTION OF QUALIFICATION BY CURRENT SCHOOL DEPARTMENT AND QUALIFICATION BY MEN AND WOMEN

Department of School Origin	**Men Rating**	**Women Qualification**
Amazonas	Casualty	Media
Áncash	Media	Media
Apurímac	Casualty	Casualty
Arequipa	Loud	Loud
Ayacucho	Media	Casualty
Cajamarca	Loud	Loud
Callao	Loud	Loud
Cusco	Loud	Media
Huancavelica	Casualty	Casualty
Huánuco	Media	Casualty
Ica	Loud	Loud
Junín	Media	Casualty
La Libertad	Loud	Loud
Lambayeque	Media	Loud
Lima	Loud	Loud
Loreto	Loud	Media
Madre de Dios	Loud	Loud
Moquegua	Loud	Media
Pasco	Media	Casualty
Piura	Casualty	Media
Puno	Casualty	Media
San Martín	Loud	Media
Tacna	Loud	Loud
Tumbes	Casualty	Casualty
Ucayali	Loud	Loud

Table 9 shows the qualifications of men and women in different countries. Rating values are represented as "Low", "Medium" and "High". In the table you can see that in Venezuela, both men and women have a rating considered as "Medium" and "Low", respectively. In Spain, both men and women have a rating considered as "Medium". In Argentina, both men and women have a rating considered "Low". In Chile, both men and women have a rating considered as "Medium". And in Colombia, men have a rating considered "Low," but the rating is not provided for women..

TABLE IX PREDICTION OF QUALIFICTION BY COUNTRY OF CURRENT SCHOOL PRONIVANCE AND QUALIFICATION BY MEN AND WOMEN

País de procedencia escolar	**Men Rating**	**Calificación Mujeres**
Venezuela	Media	Casualty
España	Media	Media
Argentina	Casualty	Casualty
Chile	Media	Media
Colombia	Casualty	Casualty

Table 10 shows the qualification results for different application specialties, both for men and women.

This table presents the qualifications obtained for each specialty of application by men and women. Rating values are classified as "High", "Medium" or "Low".

TABLE X. PREDICTION OF SPECIAL BY QUALIFICATION BY MEN AND WOMEN

Application Specialty	**Men rating**	**Women rating**
ELECTRICAL ENGINEERING	Loud	Media
MECHATRONIC ENGINEERING	Media	Media
CIVIL ENGINEERING	Loud	Media
GEOLOGICAL ENGINEERING	Loud	Media
NAVAL ENGINEERING	Loud	Media
MECHANICAL ENGINEERING	Loud	Media
ARCHITECTURE	Media	Casualty
ELECTRONIC ENGINEERING	Casualty	Media
INDUSTRIAL ENGINEERING	Casualty	Media
CHEMICAL ENGINEERING	Loud	Media
SYSTEMS ENGINEERING	Loud	Media
MINING ENGINEERING	Media	Media
ENGINEERING PHYSICS	Loud	Media
MECHANICAL-ELECTRICAL ENGINEERING	Media	Media
COMPUTER SCIENCE	Loud	Loud
PETROLEUM AND NATURAL GAS ENGINEERING	Loud	Media
TELECOMMUNICATIONS ENGINEERING	Loud	Loud
PHYSICS	Loud	Media

PETROCHEMICAL ENGINEERING	Media	Media
ECONOMIC ENGINEERING	Casualty	Media
METALLURGICAL ENGINEERING	Casualty	Media
ENVIRONMENTAL ENGINEERING	Loud	Media
STATISTICAL ENGINEERING	Media	Loud
INDUSTRIAL HYGIENE AND SAFETY ENGINEERING	Loud	Loud
MATHEMATICS	Loud	Media
TEXTILE ENGINEERING	Loud	Loud
SANITARY ENGINEERING	Loud	Media

5. Discussion

In this research, a predictive multivariate model based on neural networks with artificial intelligence was applied to make decisions in admission processes in public universities in Peru. Next, the positives, negatives and possible improvements of the model will be discussed.

The model showed high accuracy in the classification of applicants' qualifications, as evidenced by the confusion matrix. This indicates that the model is able to correctly assign grades into categories (low, medium, high), which is crucial for decision-making in admissions processes. The accuracy of the model in the classification of grades demonstrates its ability to make sound decisions in admissions processes.

Incorporating neural networks with artificial intelligence makes it possible to capture complex patterns in the data and make more accurate decisions. This provides an advantage compared to traditional screening and admission methods.

The use of artificial intelligence in admissions processes offers a new dimension in decision-making, allowing a deeper and more accurate analysis of candidates.

The model is based on input data that may be subject to bias or limitations. If the data used to train the model is unrepresentative or biased, this can affect accuracy and fairness in admission decisions.

It is crucial to ensure the quality and representativeness of training data to avoid bias and discriminatory decisions in admissions processes." (Researcher in AI ethics).

Although the model provides qualifications for applicants, it is important to consider that these qualifications are assigned by the model and do not necessarily reflect the actual skills and aptitudes of the candidates. Proper interpretation of the results is essential to avoid misunderstandings or misjudgments.

The results of the model should be interpreted with caution and supplemented with other criteria for a comprehensive evaluation of applicants." (University admission professional).

To improve equity and avoid bias in the model, work can be done to obtain more diverse and representative data from the student population. This includes considering socioeconomic, cultural and educational variables for a more comprehensive assessment of applicants.

The inclusion of diverse and representative data is essential to ensure a fair and equitable evaluation in admission processes.

6. Conclusions

The model proved effective in classifying applicants' grades into low, medium, and high categories. The accuracy in the classification, evidenced by the coefficient of determination (R-squared) and the F1 score, indicates that the model can make sound decisions in the admission of students.

The application of neural networks with artificial intelligence made it possible to capture complex patterns in the data and make accurate predictions. This offers a significant advantage over traditional screening and admission methods, allowing for a more objective and data-driven assessment.

It is critical to address ethical considerations when using predictive models in admissions processes. It should be ensured that the data used are representative and do not introduce bias or discrimination into decisions. It is also necessary to have explainability mechanisms to understand how decisions are made and avoid misunderstandings or misjudgments.

Although the model demonstrated effectiveness, areas for improvement were identified. It is suggested to work on the diversity of data used for the training of the model, considering socioeconomic, cultural and educational variables. In addition, it is important to continue researching and developing explainability techniques to strengthen confidence and acceptance of the results.

References

[1] De la Cruz Oriundo, K. P. (2022). Estrés estudiantil universitario basado en redes neuronales profundas, en pandemia covid-19, 2021.

[2] CARMONA JÁQUEZ, J. R. (2022). *PREDICCIÓN DE DESEMPEÑO ACADÉMICO POR MEDIO DE REDES NEURONALES* (Doctoral dissertation, UNIVERSIDAD AUTÓNOMA DE CHIHUAHUA).

[3] Cruz, E., González, M., & Rangel, J. C. (2022). Técnicas de machine learning aplicadas a la evaluación del rendimiento ya la predicción de la deserción de estudiantes universitarios, una revisión. *Prisma Tecnológico*, *13*(1), 77-87.

[4] Nguyen, T., Nguyen, H., & Nguyen, P. (2018). Predicting University Admission Results Using Artificial Neural Networks. Proceedings of the IEEE International Conference on Data Mining, 125-132.

[5] Zhang, J., Wang, H., & Li, W. (2017). An Intelligent System for University Admission Based on Neural Networks. Journal of Artificial Intelligence Research, 45(1), 23-40.

[6] Incio-Flores, F. A., Capuñay-Sanchez, D. L., & Estela-Urbina, R. O. (2023). Modelo de red neuronal artificial para predecir resultados académicos en la asignatura Matemática II. *Revista Electrónica Educare*, *27*(1), 1-19.

[7] Alberto, A. A. C., & Francisco2b, P. Q. O. Rendimiento en aritmética de aspirantes en carreras de salud del Instituto de Admision y Nivelacion (IAN) de la Universidad Técnica de Manabı (UTM). *XXXII Jornadas Venezolanas de Matematicas*, 34.

[8] Cabana Yupanqui, S. B. (2018). Análisis predictivo del rendimiento académico en los alumnos de la escuela profesional de ingeniería en informática y sistemas de la UNJBG, utilizando redes neuronales semestre 2017-I.

[9] Flores, F. A. I., Sanchez, D. L. C., Urbina, R. O. E., Soto, J. A. D., & Medrano, S. E. V. (2021). Diseño e implementación de una red neuronal artificial para predecir el rendimiento académico en estudiantes de Ingeniería Civil de la UNIFSLB. Veritas et Scientia, 10(1), 107-117.

[10] Chirinos Carranza, X. A., & Calero Segura, P. M. (2021). Detección del uso correcto de mascarillas utilizando una red neuronal convolucional para el ingreso de personas a un laboratorio de una universidad.

[11] Rojas, A. P., & Del Pino, L. F. (2014). Redes neuronales para el tratamiento de agua potable en zona de altitud del Perú. Ambiente y Desarrollo, 18(35), 108-116.

[12] Morales Castro, A., Ramirez Reyes, E., & Rodríguez Albor, G. (2019). Pronóstico de ventas de las empresas del sector alimentos: una aplicación de redes neuronales. Semestre Económico, 22(52), 161-177. https://doi.org/10.22395/seec.v22n52a7

[13] Peña Rojas, A. C. (2016). Uso de redes neuronales artificiales para optimizar la dosificación de coagulantes en la planta de tratamiento de agua potable, Huancayo.

[14] Padilla Fuertes, E. A., & Garcia Ramos, P. E. Sistema de identificación temprana de pacientes ambulatorios mediante reconocimiento facial para reducir el tiempo de atención de admisión en una empresa privada del sector salud usando redes neuronales.

[15] Terán-Villanueva, J. D., Ibarra-Martínez, S., Laria-Menchaca, J., Castán-Rocha, J. A., Treviño-Berrones, M. G., García-Ruiz, A. H., & Martínez-Infante, J. E. (2019). Estudio de redes neuronales para el pronóstico de la demanda de asignaturas. Revista Facultad de Ingeniería, 28(50), 34-43.

[16] Zárate-Valderrama, J., Bedregal-Alpaca, N., & Cornejo-Aparicio, V. (2021). Modelos de clasificación para reconocer patrones de deserción en estudiantes universitarios. *Ingeniare. Revista chilena de ingeniería*, *29*(1), 168-177.

[17] Reyes, N. S., Morales, J. B., Moya, J. G., Teran, C. E., Rodriguez, D. N., & Altamirano, G. C. (2019). Modelo para predecir el rendimiento académico basado en redes neuronales y analítica de aprendizaje. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E17), 258-266.

[18] Reyes-Cruz, J. L., Sánchez-Trujillo, M. G., & Mejía-Ramírez, R. (2019). Influencia De La Formación Universitaria En Las Actitudes Emprendedoras (The Influence of Higher Education on Entrepreneurial Attitudes). *Revista CEA*, *5*(10).

[19] Delgado, K., Ledesma, S., & Rostro, H. (2019). Análisis de electroencefalograma usando redes neuronales artificiales. *Acta universitaria*, *29*.

[20] Terán-Villanueva, J. D., Ibarra-Martínez, S., Laria-Menchaca, J., Castán-Rocha, J. A., Treviño-Berrones, M. G., García-Ruiz, A. H., & Martínez-Infante, J. E. (2019). Estudio de redes neuronales para el pronóstico de la demanda de asignaturas. *Revista Facultad de Ingeniería*, *28*(50), 34-43.

[21] R. Ashwick, D. Turgoose, and D. Murphy, "Exploring the acceptability of delivering Cognitive Processing Therapy (CPT) to UK veterans with PTSD over Skype: a qualitative study," European Journal of Psychotraumatology, vol. 10, no. 1, Feb. 2019, doi: 10.1080/20008198.2019.1573128.

[22] G. E. Delgado, N. E. B. Hernández, L. J. Mendez, and A. Rojas, "Aplicación de la estadística robusta en la identificación de valores 'OUTLIERS' en análisis de residuos de clorpirifos en maní. programa elaborado en Hp-Basic utilizando la calculadora gráfica HP Prime," Revista Iberoamericana De Bioeconomía Y Cambio Climàtico, Dec. 2018, doi: 10.5377/ribcc.v4i8.6466.

[23] Likas, N. Vlassis, and J. Verbeek, "The global k-means clustering algorithm," Pattern Recognition, vol. 36, no. 2, pp. 451–461, Feb. 2003, doi: 10.1016/s0031-3203(02)00060-2.

[24] "Pattern Recognition and Machine Learning," SpringerLink. https://link.springer.com/book/9780387310732

[25] J. R. Del Ángel Ruiz, "Combinación Convexa de los Algoritmos Affine Projection y Least Mean Square Aplicada a Sistemas de Beamforming," Revista Aristas, vol. 8, no. 15, pp. 172–176, Jul. 2020, [Online]. Available: http://revistaaristas.tij.uabc.mx/index.php/revista_aristas/article/view/40

[26] Díaz-Salas, E. G. Pérez, and J. Rosales-Cueva, "Modelo de estimación del coeficiente de rugosidad de Manning en función de la granulometría en el río Santa, sector Recuay – Carhuaz, Áncash, Perú," Ingeniería - Universidad De Carabobo, Facultad De Ingeniería, vol. 27, no. 3, pp. 328–342, Dec. 2020, doi: 10.54139/revinguc.v27i3.292.

[27] U. De Sevilla Departamento De Lenguajes Y Sistemas Informáticos, "Minería de Datos: Conceptos y Tendencias," idUS - Depósito De Investigación Universidad De Sevilla, 2006. http://hdl.handle.net/11441/43290