Web Application Based on Neural Networks for the Detection of Students at Risk of Academic Desertion

Manuel S. Asto-Lazaro [1], Segundo E. Cieza-Mostacero [1]

[1] *Research Group Trend and Innovation in Systems Engineering-Trujillo, Universidad César Vallejo, Avenue Larco 1770 Trujillo, Peru*

***Abstract* – The objective of the study is to improve the prediction of academic desertion by means of a web application based on neural networks in a private university in Trujillo during 2023 The study employed and applied pure experimental design. The population comprised all the faculty's academic processes, with a sample size of 60 processes. The data collection technique was direct observation, and the instrument was an observation sheet. The web application was developed using Python 3 and the Flask framework, with MySQL as the database manager and Extreme Programming (XP) as the software development methodology. For the descriptive analysis, Microsoft Excel 2019 was used, while Jamovi 2.3.28 was utilized for inferential analysis. The results showed a reduction of 123 seconds in the average time for academic data collection, a reduction of 848 seconds in the average prediction time for identifying students at risk of academic dropout, and a 3% increase in the identification rate of at-risk students. In conclusion, the use of a web application based on neural networks significantly improves the prediction of academic dropout.**

DOI: 10.18421/TEM133-83
https://doi.org/10.18421/TEM133-83

Corresponding author: Manuel S. Asto-Lazaro,
Research Group Trend and Innovation in Systems Engineering-Trujillo, Universidad César Vallejo, Avenue Larco 1770 Trujillo, Peru
Email: mastola23@ucvvirtual.edu.pe

Received: 07 February 2024.
Revised: 11 June 2024.
Accepted: 05 August 2024.
Published: 27 August 2024.

***Keywords* – Web application, academic desertion, machine learning, artificial neural networks, prediction.**

1. Introduction

Academic desertion prevents the achievement of higher educational levels and directly affects students, and can be caused by various academic, economic, social, demographic, emotional, and other factors. Therefore, it is essential to adopt the necessary measures to ensure that the highest percentage of students complete their academic development and therefore can achieve better opportunities in the labor market [1].

Normally during the first semesters, students are confronted with the reality of higher education, challenging their expectations, which makes them more prone to a variety of drawbacks of different origins. Such as financial factors, lack of information and vocational orientation, as well as prior education, all of which can have a significant influence on the abandonment of their academic education [1].

Between 2000 and 2021, Colombia reported an annual dropout rate of 11%, in other words, 1 in 10 students dropped out of higher education. In addition, during 2020, the year when the COVID-19 pandemic began, a dropout rate of 12.7% was reported, which is higher than the average [1]. On the other hand, in Peru during 2020, the university dropout rate was higher than average compared to 2018 and 2019 [2].

After the inquiries it was noted that there are problems that are manifested in a private university in Trujillo, regarding the prediction of academic desertion: The department of university welfare states that there is a delay in the time it takes to collect academic data on enrolled students, because there are several individual reports and the lack of a consolidated report requires reviewing each one separately, leading to the expenditure of more time for the collection of information.

The university welfare department also states that there is a delay in the time it takes to predict students at risk of academic desertion, since the review of attendance and analysis of academic data is done manually, resulting in the lack of early intervention to prevent desertion. Likewise, the administrative staff states that the percentage of students at risk of academic desertion is unknown, due to the fact that there is an outdated report on dropouts, and that the number of students at risk of academic desertion is not quantified and analyzed, resulting in the lack of retention strategies and the failure to make appropriate decisions to prevent academic desertion.

Therefore, the general research problem was formulated as: How does a web application based on neural networks influence the prediction of academic desertion in a private university in Trujillo during the 2023 period? Likewise, the specific problems were how does a web application based on neural networks reduce the average time of academic data collection? How does a web application based on neural networks reduce the average prediction time of students at risk of academic desertion? How does a web application based on neural networks increase the percentage of identified cases of students at risk of academic desertion?

Finally, it was established as the general hypothesis that if a web application based on neural networks is used, then the prediction of academic desertion in a private university in Trujillo will be improved, and as a specific hypothesis that if a web application based on neural networks is used, then the average time of academic data collection will be reduced, if a web application based on neural networks is used, then the average prediction time of students at risk of academic desertion is reduced, and if a web application based on neural networks is used, then the percentage of identified cases of students at risk of academic desertion is increased.

2. Literature Review

Next, the theoretical bases used in the research are mentioned, which are related to the dependent variable (Prediction of Academic Dropout) and the independent variable (Web Application Based on Neural Networks).

2.1. Academic Desertion

Implies the indefinite abandonment of an academic program by the student, due to economic, sociological, psychological, organizational factors, prior knowledge, and academic performance, among others [3]. In addition, there are types of desertion such as early dropout, when the student leaves school before enrolling.

Initial, when the student leaves school in the course of the first cycles, and late dropout, when the student leaves school after the fifth cycle [4].

2.2. Machine Learning

Machine learning is a part of artificial intelligence (AI) that involves the creation of systems with the ability to learn on their own, from a dataset [5]. It also comprises a variety of algorithms so that machines are able to learn and provide solutions to highly complex problems [6].

In the development of machine learning projects, three fundamental phases are identified. The first phase covers the collection and processing of data, in this stage the main objective is to obtain the necessary information to analyze and carry out correction or modification operations of the original data, if necessary. The second phase consists of storing the information obtained for later analysis. The third phase involves the application of machine learning techniques or algorithms with the aim of obtaining a predictive model and/or just the extraction of knowledge [7].

2.3. Artificial Neural Networks

Among the various techniques and algorithms of machine learning are artificial neural networks, which are models inspired by the human brain. These networks are made up of interconnected artificial neurons, aiming to function similarly to biological neurons, being able to imitate the behavior of the brain during the learning process [6]. Using a set of operations, they adjust their behavior to obtain a result against an initial data entry. They are able to recognize patterns and make approximations [8].

2.4. Data Pre-Processing

In most cases, the datasets present a lack or inconsistency of the data due to various causes, such as the absence of information at the time of providing it to errors during its capture. If the problem is not properly addressed, it can lead to errors or unintended results. That is why data preprocessing is essential in machine learning processes. There are techniques for handling missing data, such as removing or imputing missing values. In addition, because most machine learning algorithms are conditioned to work with numerical values, it is necessary to transform categorical data into numerical values using a variable coding technique. Finally, scaling is another important process of data preprocessing, because machine learning algorithms work best if the characteristics of the dataset are at the same scale [8].

2.5. Model Evaluation

Regarding the evaluation of classification models, Bosch *et al*. [9] mention that the performance of such models is calculated by comparing the predictions obtained with respect to the true cases present in the dataset used. He also mentions that the confusion matrix shows a graphical overview of the errors and successes obtained by the predictive model, it is also known as the error matrix or contingency matrix. From this matrix, a series of metrics are identified that allow the performance of the model to be quantified.

3. Related Works

Some similar works on the prediction of academic desertion by means of predictive models are mentioned below. These investigations were extracted from various digital databases such as EBSCO, Scopus, Proquest, Scielo, etc.

Guerra *et al.* [10] evaluated 5 machine learning algorithms, with the aim of developing a predictive model of academic desertion, such algorithms evaluated were: artificial neural networks, k-nearest neighbors (KNN), decision trees, support vector machines (SVM), and random forest. In addition, they relied on the KDD (knowledge discovery database) methodology. They also used online surveys and questionnaires as data collection techniques and tools. The Scikit-learn library was also used for model generation. In addition, as a result, a model with an accuracy of 0.92 was obtained, using the artificial neural network algorithm, while with the other algorithms an accuracy below 0.86 was obtained. Finally, it was concluded that the artificial neural network algorithm allowed the creation of a model with better performance in the prediction of dropout.

Zapata [11] also aimed at detecting students at risk of dropping out by using a method based on the design of metrics and educational data mining techniques. He had a dataset of 10228 records on students of secondary education levels in Colombia. It also obtained a sample of 4101 records. The researcher also obtained a model with an accuracy of 0.88 and sensitivity of 0.65 using metric-based feature transformation, while the model obtained an accuracy of 0.84 and sensitivity of 0.57 using metric-free features. Finally, it was concluded that there was an increase in the performance of the predictive model using the proposed feature transformation method.

Lastly, Smith and Gutiérrez [12] aimed to determine whether machine learning allowed a more accurate early warning compared to the logistic regression method. They relied on a dataset of 3 399 147 records and 21 variables, coming from a database provided by the Chilean Ministry of Education. In addition, Microsoft SQL Server 2017 was used for data processing and Microsoft R for statistical analysis. Artificial neural networks, decision trees, and random forest were used as machine learning algorithms. Likewise, a sensitivity of 28.65% was obtained using logistic regression, whereas sensitivity greater than 39% was achieved using models generated from machine learning algorithms. In addition, both models achieved an accuracy greater than 90%. As a result of the higher sensitivity obtained by applying machine learning algorithms, it was concluded that they allowed a higher detection percentage of dropout cases.

4. Purpose of the Study

As a general objective, it was determined to improve the prediction of academic desertion by means of a web application based on neural networks in a private university in Trujillo during the period of 2023. In addition, the specific objectives are also mentioned: to reduce the average time of academic data collection, to reduce the average prediction time of students at risk of academic desertion and finally to increase the percentage of identified cases of students at risk of academic desertion.

5. Methodology

The research was of an applied type, since it sought to solve existing practical problems through the application of theory, in addition to the fact that this type of research uses empirical experiments for the generation of data in a field of study. As a pure experimental research design, since we sought to evaluate the effects of the independent variable on the dependent variable, using a randomly selected control group and an experimental group [13].

The independent variable was a web application based on neural networks and is defined as a computer application that uses artificial neural network algorithms and a client-server architecture to process data and offer services through the Internet [14]. This variable will be measured with the verification of compliance (presence) or not (absence) of the implementation of the web application based on neural networks, for this the presence/absence indicator considered within the nominal measurement scale was taken into account.

The dependent variable was the prediction of academic desertion and is defined as the process that involves using statistical or machine learning models to identify in advance the students with the highest risk of dropping out of school before completing an academic program [15].

For this variable, the observation sheet tool was used and measured through the following indicators: Average time of academic data collection (ATADC), average prediction time of students at risk of academic desertion (APTSRD) and percentage of identified cases of students at risk of academic desertion (PICSRD). Such indicators were considered within the ratio measurement scale.

In addition, all the processes of predicting academic dropout in the Private Universities of Peru were considered as population. For the sample of the present research, 30 prediction processes were taken into account. In addition, simple random probability sampling was considered, since it involves the random selection of the elements that make up the population.

Likewise, direct observation was considered as a data collection technique, since it is characterized by the fact that the researcher observes the object of study and obtains the information directly from the population studied, allowing the behavior of the phenomenon in question to be observed. In addition, the observation sheet was considered as a data collection instrument, since it is used to analyze or evaluate the object of study [13].

6. Development of the Proposal

The proposal presents a web application, which allows the user to detect students at risk of academic desertion, this website was based on a client-server architecture. In addition, it is developed with Flask, HTML, CSS and JavaScript.

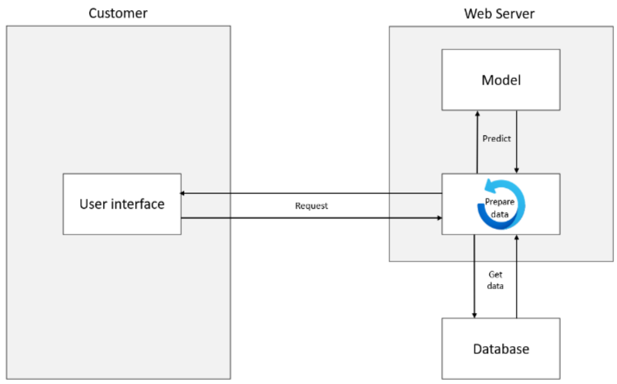

Figure 1. Web application client-server architecture

6.1. Data Collection

During this stage, data was gathered from a computer system belonging to a private university in Trujillo, Peru, and an exploratory analysis was performed in order to better understand the characteristics and structure of the dataset obtained [16].

Since the information system had a large amount of data, only tables and columns related to undergraduate students entering in semesters 2019-1 to 2022-1 were taken under consideration, so a dataset with 6312 records and 18 columns was obtained in the first instance (Table 1).

Table 1. Description of gathered dataset

N°	Attribute	Description	Type
1	GENDER	Student Gender.	Text
2	DISABILITY	Whether the student has a disability.	Binary
3	EMPLOYED	Whether the student is employed.	Binary
4	MARITAL_STATUS	Student’s marital status.	Text
5	EDUCATIONAL_LEVEL	Student´s highest level of instruction.	Text
6	AGE	Student´s age.	Numerical
7	ACCEPTANCE_TYPE	Student’s Acceptance type.	Text
8	FACULTY	Student’s faculty.	Text
9	EDUCATIONAL_MODALITY	Student’s Educational Modality.	Text
10	SEMESTER	Semester in which the student is currently enrolled.	Numerical
11	PAYMENT_CATEGORY	Student’s sliding scale category.	Text
12	PCA	Student’s percentage of approved academic credits.	Numerical
13	PPA	Student’s academic weighted average.	Numerical
14	PSA	Student’s current semester academic average.	Numerical
15	PAA	Student’s approved subject average.	Numerical
16	INTERN	Whether the student is also an intern.	Binary
17	ABSENCE	Student’s absences percentage.	Numerical
18	PCP	Student’s percentage of pending installments.	Numerical

As part of the exploratory analysis, it was found that the faculty of humanities had a higher dropout rate compared to the other faculties, trailed by the faculty of management and economics (Table 2).

In addition, from the dataset obtained, several classes were identified referring to the student's status (freshman, incomplete, graduate, etc.), from which only two were taken under consideration for the research (dropouts and non-dropouts).

Table 2. Desertion by faculty freshmen of semesters 2019-1 – 2022-1

N°	Faculty	% Desertion
1	Management And Economics	16
2	Health Science	4
3	Law And Political Sciences	6
4	Humanities	63
5	Engineering And Architecture	10
6	Teology	1

6.2. Feature Gathering

In machine learning, feature selection is an indispensable process to improve model quality, reduce complexity when there are many variables, reduce storage space, and reduce preprocessing and training time. Thus, feature selection techniques enable the identification of the most relevant features [17].

This, during this stage, a random forest algorithm was applied to estimate the importance of the features provided by the dataset, using SelectFromModel from the sklearn library. In addition, the most relevant features were identified (Table 3).

Table 3. Importance of the upmost significant variables

N°	Feature	Importance
1	PCP	0.19004568
2	PSA	0.14962306
3	PPA	0.13204959
4	ABSENCE	0.12000246
5	AGE	0.08691038
6	PAA	0.07610645
7	PCA	0.06660278

In addition, since the data used is from undergraduate students, it was noticed that the EDUCATIONAL_LEVEL of all the records was completed secondary education, so it was decided to discard this characteristic, as was the case with PAYMENT_CATEGORY and INTERN, since no different values were present, it was decided to discard them all [18]. Table 4 shows the discarded variables.

Table 4. Discarded variables

N°	Attribute	Type
1	EDUCATIONAL_LEVEL	Text
2	ACCEPTANCE_TYPE	Text
3	FACULTY	Text
4	PAYMENT_CATEGORY	Text
5	INTERN	Binary

6.3. Dataset Preparation

Since the dataset contained null values, data cleaning was performed in order to obtain a more effective model [12]. Then, selection filters were applied, due to the existence of uniform records. Subsequently, class balancing was applied to adjust the distribution of the data set, because the number of records on non-dropout students was greater than the number of dropout students, causing the model to be biased towards the predominant class. As a result, a new dataset with 5050 records was obtained. Finally, the categorical variables were coded using the OneHotEncoder technique of the sklearn library (see Table 5).

Table 5. Categorical variables identified on the dataset

N°	Attribute	Type
1	GENDER	Text
2	MARITAL_STATUS	Text
3	EDUCATIONAL_MODALITY	Text

6.4. Training

In this phase, 20% of the dataset was taken for the testing data and the remaining 80% was taken for the training data. Then we proceeded to normalize the values of the input variables in order to ensure that they all have the same scale, using StandardScaler from the sklearn library.

Starting with the training, the X value was taken as the dataset features and the Y value as the labels to be predicted (dropout or non-dropout). Then we proceeded with the creation of the model, which consists of 1 input layer, 2 hidden layers, and an output layer, the architecture and hyperparameters used are shown below. A threshold of 0.7 was also established, i.e. the predictions made by the model with a value greater than 0.7 are considered as 1 (deserters). The hyperparameters for model training were also established using the fit function of the tensorflow library (Table 6).

Table 6. Model architecture

Hyperparameter	Input Layer	Hidden Layer 1	Hidden Layer 2	Output Layer
Neurons	32	32	32	1
Weight Indexing	Uniform			
Activation Function	ReLU			Sigmoid
Input Dimensions	64	-	-	-
Dropout Percentage	-	0.1	0.1	-
Optimizer	Adam			
Loose Function	Binary Cross entropy			
Metrics	Accuracy			
Training Parameters				
Hyperparameter	Value			
batch_size	10			
epochs	100			

The confusion matrix is a tool to assess performance of a predictive model and provides a view of the predicted output compared to the actual dataset values. TP (True positive) means that the model was correct in predicting the output as positive, FN (False negative) shows that the model was wrong in predicting the output as negative, FP (False positive) indicates that the model was wrong in predicting the output as positive and TN (True negative) refers that the model was correct in predicting the output as negative [8] (Table 7).

Table 7. Confusion matrix obtained during model training

	DESERTOR (PP)	NON DESERTOR (PN)
DESERTOR (TP)	TP (286)	FN (45)
NON DESERTOR (TN)	FP (122)	TN (522)

Based on the confusion matrix, performance metrics can be obtained such as: accuracy, which corresponds to the ratio between correct predictions and the total number of predictions made; precision, which consists of the proportion of correct positive predictions in relation to the total number of positive predictions; recall, which is the number of elements correctly identified as positive out of the total number of true positives; and F1-score, which is the harmonic mean of the precision and recall of a model [19]. Therefore, the aforementioned metrics were calculated to assess the performance of the generated model (Table 8).

Table 8. Performance metrics of the obtained model

Algorithm	Accuracy	Precision	Recall	F1 Score
Artificial Neural Network	0.81	0.88	0.72	0.79

6.5. Backend and Frontend Development

For the development of the web application, the extreme programming (XP) methodology was used, which is an agile development framework that allows the development of quality software. Some of its features are that it is oriented around the development team and, moreover, it promotes early user any client participation. It is based on 5 principles: communication, simplicity, feedback, courage, and respect [20].

The phases of the extreme programming (XP) development methodology begin with planning, where the user stories are created and the requirements are specified; then the design phase continues, in this stage an operational prototype of the software is made; then comes the implementation phase, where the software coding and implementation of the user stories begins; and finally ends with the testing phase, where the software is tested and validated according to the user stories created in the planning [21].

6.6. Phase 1: Planification

In this phase the project roles were assigned, such as programmer, tester, and big boss. Also, 4 user stories were created, which contained the requirements mentioned by the users with respect to the application. The revision schedule for each story was also established (Table 9).

Table 9. Project progress review timeline

Story	Story Name	Priority	Date
H001	Prediction of students at possible risk of desertion	HIGH	2023-09-18
H002	Integration with ERP database	HIGH	2023-09-10
H003	System Access	HIGH	2023-09-21
H004	Results reporting	HIGH	2023-09-25

6.7. Phase 2: Design

In this phase the user interface prototypes were defined, CRC cards (class, responsibility, and collaboration) and the database structure were developed.

The database was developed using MySQL and 7 tables were defined: predictions, where the values predicted by the application are stored; users, where data of the users accessing the application are stored; models, used to record information of the trained model; trained, where the history of the metrics obtained during the training of a model is stored; and parameters, where static values used by the application are stored.

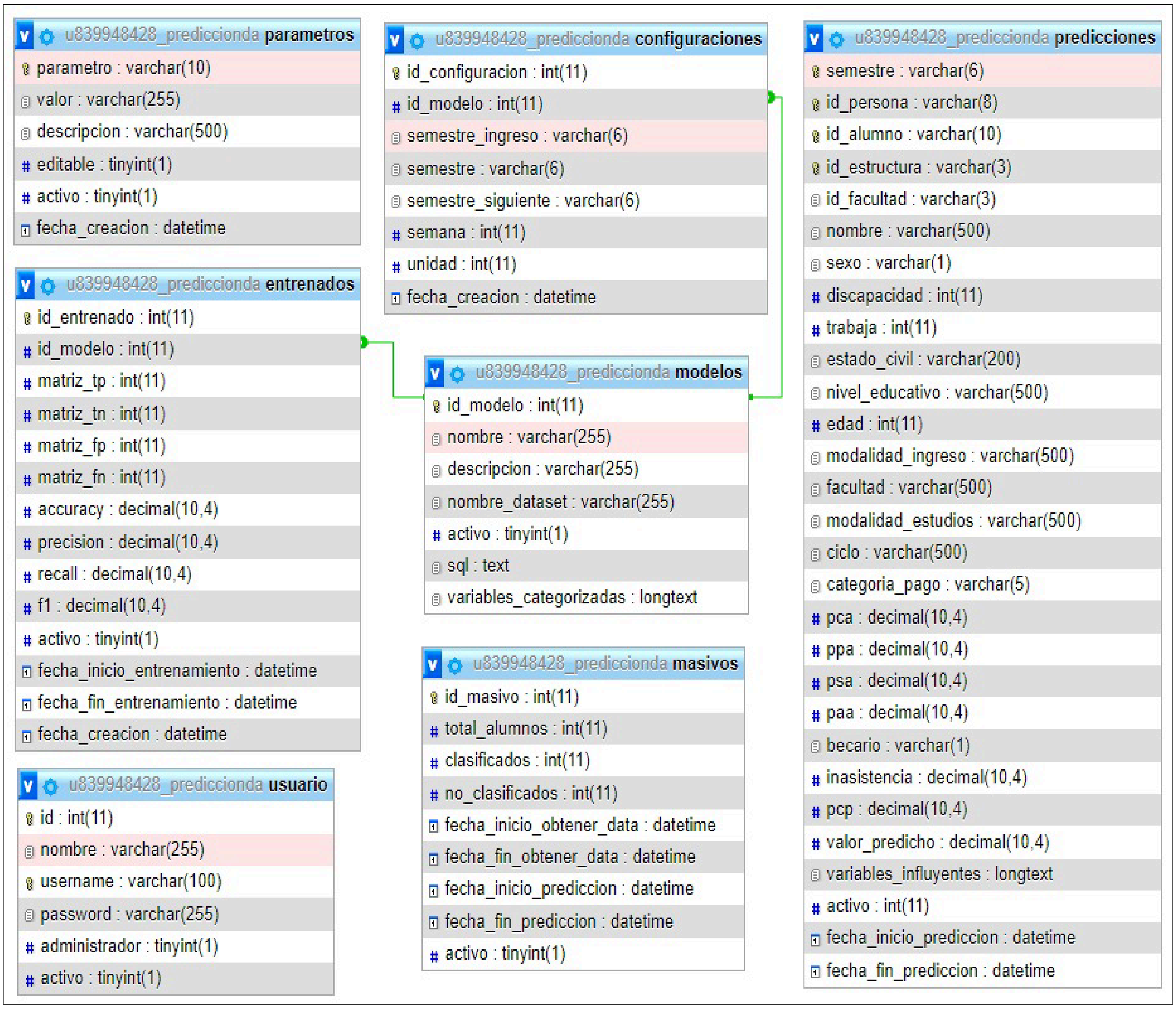

Figure 2. Database diagram

6.8. Phase 3: Coding

Flask was used for backend development, which is a microframework, made in Python that offers a simple way to develop web applications but potentially scalable. In addition, it has good documentation and a large active community [10].

After training the model, we proceeded to export it with the "keras" extension and created a function that allows us to load the model, and then provide it with new data to obtain new predictions. In addition, since the dataset contained categorical attributes, it was necessary to store the categorized values for later reuse, and the results obtained from each prediction were saved in a database created with MySQL.

Dashboard: After a successful login, an interface is displayed that includes a traffic light with the number of students at risk of dropping out detected by the application, classified into low, medium, and high levels. To identify each level, ranges were established as follows: low [0, <= 0.33], medium [> 0.33, <= 0.70] and high [> 0.70]. In addition, a graph showing the history of students at high risk of dropping out per semester is also displayed.

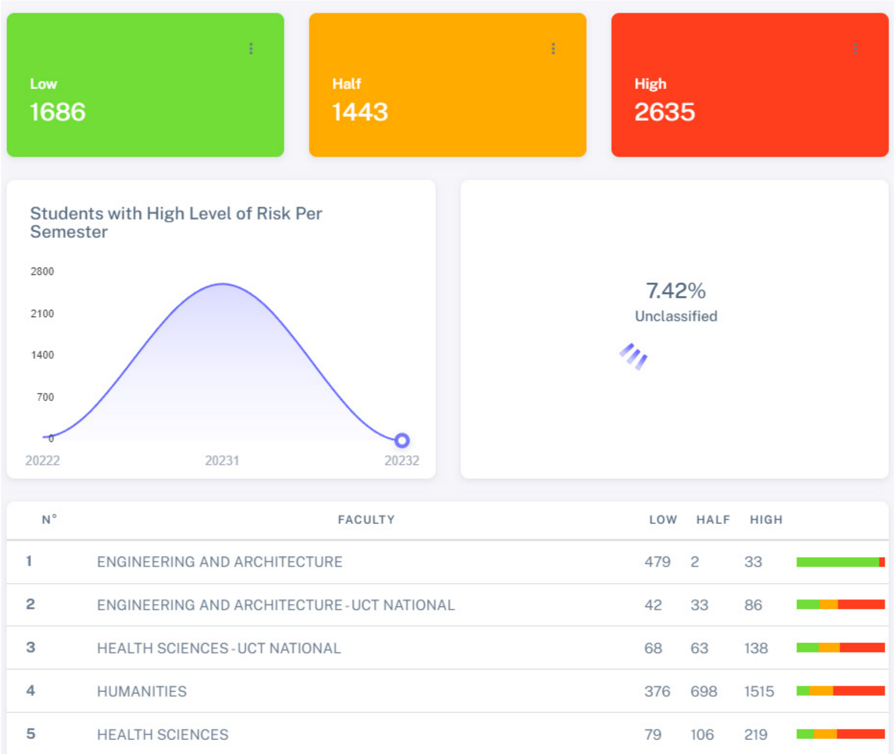

Figure 3. Dashboard interface

Search for courses and sections: The application displays an interface that prompts users to select the semester and contains a search bar for teachers by identity document number or name. In addition, it contains a "Search Courses" button that allows consulting the courses taught by a teacher. Additionally, once the query has been made, the list of courses and sections is displayed with a "View" button that allows users to consult the list of students enrolled in a section.

Prediction by section: The application shows an interface with the data of the professor, career, class and the list of students of the previously selected section, it also shows the percentages of dropout risk for each student, Also, it includes a "Process" button so that new predictions can be generated.

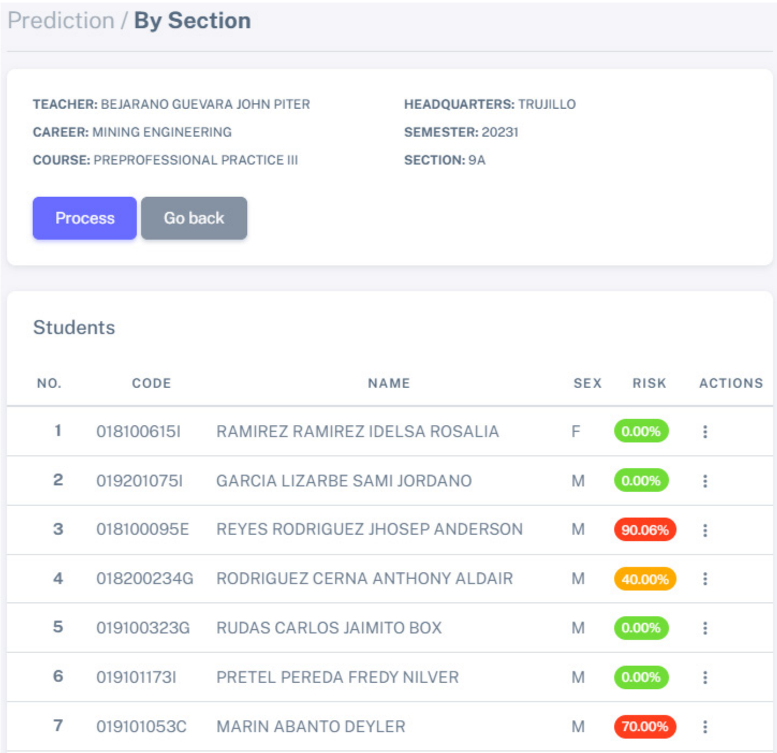

Figure 4. Prediction interface by section

To make new predictions, a method named "predict" was implemented, which receives the values of the input variables (x) and the name of the model to be used, then a file with extension “pkl” is loaded containing a StandardScaler object with the settings used in the model training, then the x data is transformed using the StandardScaler object, then the model previously exported with the “keras” extension is loaded, then the "predict" method is used to obtain a new prediction and finally the value obtained is returned.

```python
def predict(x, model_name):
    path = config.PATH_ML + model_name + '/'

    # Loading the Adjusted StandardScaler Object
    sx = joblib.load(path + model_name + '.pkl')

    # Standardizing data
    x = sx.transform(x)

    # Loading the model
    model = tf.keras.models.load_model(path + model_name + ".keras")

    # Making Prediction
    result = model.predict(x)
    return result[0][0]
```

Figure 5. Source code of "predict" method

Results report: The application displays an interface with the list of students at risk of desertion previously detected and the predicted percentages. In addition, it contains a filter by faculty and career, and also includes an option in the "Actions" column to visualize student data.

7. Results

Next, we describe the descriptive analysis and inferential analysis about the results obtained during the research on the improvement of the prediction of academic dropout in a private university of Trujillo through a web application based on neural networks.

7.1. Descriptive Analysis

The post-test values of the control group (CG) and experimental group (EG) of the research indicators: average time of academic data collection (ATADC), average prediction time of students at risk of desertion (APTSRD) and percentage of identified cases of students at risk of desertion (PICSRD) are shown below.

Table 10. Post-test results for the ATADC indicator of the control and experimental groups

Research Indicators	Average		% Less than average	
	Control Group	Experimental Group	EG	CG
ATADC	132	9	56.67	100

Table 10 shows that the control group obtained 132 seconds in the average time of academic data collection, while the experimental group obtained 9 seconds, showing a decrease of 123 seconds.

Table 11. Post-test results for the APTSRD indicator of the control and experimental groups

Research Indicators	Average		% Less than average	
	Control Group	Experimental Group	EG	CG
APTSRD	1216	368	53.33	100

Table 11 shows that the control group obtained 1216 seconds in the average prediction time of students at risk of desertion, while the experimental group obtained 368 seconds, showing a decrease of 848 seconds.

Table 12. Post-test results for the PICSRD indicator of the control and experimental groups

Research Indicators	Average		% Higher than average	
	Control Group	Experimental Group	EG	CG
PICSRD	5	8	40.00	60.00

Table 12 shows that the control group obtained an average of 5% in the percentage of cases detected among students at risk of desertion, while the experimental group obtained an average of 8%, showing an increase of 3%.

7.2. Inferential Analysis

For the inferential analysis, normality test and hypothesis testing were performed, likewise, the following decision criteria were proposed for the post-test of the control group (CG) and the experimental group (EG) of the indicators: average time of academic data collection (ATADC), average time of prediction of students at risk of desertion (APTSRD), and percentage of identified cases of students at risk of desertion (PICSRD): H0: The post-test data have a normal distribution. Ha: The posttest data does not have a normal distribution. If $p < 0.05$, then the null hypothesis (H0) is rejected and the alternate hypothesis (Ha) is accepted. If $p >= 0.05$, then the null hypothesis (H0) is accepted and the alternate hypothesis (Ha) is rejected.

Since the amount of data for each indicator is under 50 entries, Shapiro-Wilk was used to perform the normality test.

Table 13. Normality test using Shapiro-Wilk for each indicator

N°	Indicator	p (CG)	p (EG)
1	ATADC	< .001	< .001
2	APTSRD	0.018	0.001
3	PICSRD	0.821	< .001

Therefore, according to the data obtained in Table 13, it is concluded that the p-value is less than 0.05 in at least one of the groups, and therefore the data does not have a normal distribution. Consequently, the nonparametric U Mann-Whitney test was applied to analyze the differences between independent groups. In addition, for hypothesis testing, μ1 was considered as the control group and μ2 as the experimental group.

Table 14. Mann-Whitney U test for each indicator

N°	Indicator	H0	Ha	p
1	ATADC	$\mu1 <= \mu2$	$\mu1 > \mu2$	0.001
2	APTSRD	$\mu1 <= \mu2$	$\mu1 > \mu2$	< .001
3	PICSRD	$\mu1 >= \mu2$	$\mu1 < \mu2$	0.048

Therefore, according to the data obtained in Table 14, there is evidence that the p-value is less than 0.05 for each of the indicators, so the null hypothesis (H0) is rejected and the alternative hypothesis (Ha) is accepted for each indicator.

8. Discussions

Regarding the indicator average time of academic data collection (ATADC) in the post-test of the control group (CG), an average time of 132 seconds was obtained, while in the post-test of the experimental group (EG) an average collection time of 9 seconds was obtained, so that a decrease of 123 seconds in the average collection time was evidenced. Likewise, Carranza [22], in his research on the influence of business intelligence on academic decision-making, obtained as part of the results a 6.7 decrease in average time in the generation of reports on approved students and a 6.8 decrease in average time in the generation of reports on failed students.

Regarding the indicator average prediction time of students at risk of academic desertion (APTSRD) in the post-test of the control group (CG), an average time of 1216 seconds was obtained, while in the post-test of the experimental group (EG) an average prediction time of 368 seconds was obtained, so that a decrease of 848 seconds in the average prediction time was evidenced. Likewise, Capuñay [23], in his research on the construction of a model based on artificial neural networks, for the projection of academic performance in high school students, obtained as part of the prediction results an average time of 20 seconds during 8 iterations.

With respect to the indicator percentage of identified cases of students at risk of academic desertion (PICSRD), in the post-test of the control group (CG) 5% of detected cases was obtained, while in the post-test of the experimental group (EG) 8% of detected cases were obtained, so that an increase of 3% of detected cases was evidenced. Likewise, Guerra *et al.* [10], in their research on the development of a predictive model of university dropout through the evaluation of machine learning algorithms, obtained an average sensitivity of 90% with the artificial neural networks algorithm.

9. Conclusion

Given that in the post-test of the control group (CG) 132 seconds was obtained as the time average for collecting academic data and 9 seconds in the post-test of the Experimental Group (EG), it was concluded that there is a decrease of 123 seconds in the time average for collecting academic data.

After obtaining, in the post-test of the control group (CG) 1216 seconds as the time average for the prediction of students at risk of desertion and 368 seconds in the post-test of the experimental group (EG), it was concluded that there is a decrease of 848 seconds as the time average for the prediction of students at risk of desertion.

Having obtained 5% of identified cases of students at risk of desertion in the post-test of the control group (CG) and 8% of identified cases in the post-test of the experimental group (EG), it was concluded that there is an increase of 3% of identified cases of students at risk of desertion.

Finally, it was concluded that, if a web application based on neural networks is used, then it improves the prediction of academic desertion in a private university in Trujillo.

References:

[1]. Laboratory of Economics of Education. (2023). *Dropout in higher education in Colombia, 74*, 17. Pontifical Javeriana University. Retrieved from: https://lee.javeriana.edu.co/-/lee-informe-74 [accessed: 01 February 2024].

[2]. Ministry of Education, General Directorate of University Higher Education, & Directorate of Policies for the Development and Quality Assurance of University Higher Education. (2021). *Report on the interruption of university studies in Peru, in the context of COVID-19*. Ministry of Education. Retrieved from: https://repositorio.minedu.gob.pe/handle/20.500.12799/7742 [accessed: 02 February 2024].

[3]. Soler, M. G., Cárdenas, F. A., Blanco, D. A., & Umbarila, X. (2020). *Student Attrition: A Cohort Research in the Bachelor of Science in Chemistry Program (1. ed.).* National Pedagogical University. Retrieved from: https://www.digitaliapublishing.com/a/70926/desercion-estudiantil--una-investigacion-de-cohorte-en-el-programa-de-licenciatura-en-quimica [accessed: 03 February 2024].

[4]. Díaz, B., Marín, W., Lioo, F., Baldeos, L., Villanueva, D., & Ausejo, J. (2022). Deserción de estudiantes, factores asociados con árboles de decisión: caso Escuela de Postgrado 1de una Universidad pública en Perú. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E53), 197-211.
Ortega, J. M. (2022). *Big data, machine learning y data science en python*. Ra-Ma Editorial. Retrieved from: https://www.digitaliapublishing.com/a/127652/big-data--machine-learning-y-data-science-en-python [accessed: 04 February 2024].

[5]. Restrepo, D. A., Viloria Porto, J. P., & Robles Algarín, C. A. (2021). *The road to artificial neural networks*. Editorial Unimagdalena.

[6]. Casas, J., Nin, J., & Julbe López, F. (2019). *Big Data: Data Analysis in Massive Environments (1st Ed.).* Editorial UOC.

[7]. Pineda, C. M. (2022). *Machine and Deep Learning in Python*. Rama Editorial. Retrieved from: https://www.digitaliapublishing.com/a/127668/aprendizaje-automatico-y-profundo-en-python [accessed: 05 February 2024].

[8]. Bosch, A., Casas, J., & Lozano, T. (2020). *Deep Learning: Principles and Fundamentals (1st ed.).* Editorial UOC.

[9]. Quishpe-Morales, S., Pillo-Guanoluisa, D., Revelo-Portilla, I., & Guerra-Torrealba, L. (2020). Modelo de predicción de la deserción universitaria mediante analítica de datos: Estrategia para la sustentabilidad. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E35), 38-47.

[10]. Zapata, D. (2021). *Method for the detection of students at risk of dropping out, based on a metrics design and a data mining technique* [Bachelor's thesis, National University of Colombia]. Retrieved from: https://repositorio.unal.edu.co/handle/unal/80615 [accessed: 05 February 2023].

[11]. Smith, J., & Gutiérrez, C. (2022). An Application of Machine Learning in Public Policy. Early Warning Prediction of School Dropout in the Chilean Public Education System. *Multidisciplinary business review*, *15*(1). Doi: 10.35692/07183992.15.1.4

[12]. Arias, J. L., Holgado, J., Tafur, T., & Vasquez, M. J. (2022). *Research methodology: The ARIAS method for developing a thesis project*. Editorial Inudi. DOI: https://doi.org/10.35622/inudi.b.016

[13]. Lozano, F. J., Alcázar, D., & Díaz, A. (2022, June). Web-based application for generation of video-interview summaries using neural networks. In *2022 17th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1-3). IEEE.

[14]. Altamirano, G. E. V. (2023). Modelo de predicción de deserción escolar en los estudiantes de la unidad educativa Los Andes por impacto de la pandemia. *Ciencia Latina Revista Científica Multidisciplinar*, *7*(1), 3038-3052. Doi: 10.37811/cl_rcm.v7i1.4640

[15]. Urcuqui, C. C., Navarro, A., Quintero, L. F., Rivera, J. A., Angulo, J. F., García, D., Sarasti, M. A., Arce, A., Samacá, A., & Torres, D. A. (2023). *Cybersecurity: Data, the seeds of chaos*. Icesi University. Retrieved from: http://repository.icesi.edu.co/biblioteca_digital/handle/10906/106425 [accessed: 06 February 2024].

[16]. Franco, E. A., Martínez, R. E. L., & Domínguez, V. H. M. (2021). Modelos predictivos de riesgo académico en carreras de computación con minería de datos educativos. *Revista de Educación a Distancia (RED)*, *21*(66). Doi: 10.6018/red.463561

[17]. Martin Lamas Piñeiro, J. J., & Wong Portillo, L. R. (2022). Web architecture for URL-based phishing detection based on Random Forest, Classification Trees, and Support Vector Machin. *Inteligencia Artificial: Revista Iberoamericana de Inteligencia Artificial*, *24*(69). Doi: 10.4114/intartif.vol25iss69pp107-121

[18]. Alhazmi, E., & Sheneamer, A. (2023). Non-Orthogonal Random Access (NORA) in Large Cellular Networks. *IEEE ACCESS*, *11*, 27579-27589. Doi: 10.1109/ACCESS.2023.3250702

[19]. Shrivastava, A., Jaggi, I., Katoch, N., Gupta, D., & Gupta, S. (2021, July). A systematic review on extreme programming. In *Journal of Physics: Conference Series*, *1969*(1), 012046. IOP Publishing. Doi: 10.1088/1742-6596/1969/1/012046

[20]. Jimenez, C. (2021). *UML. Application architecture in Java, C++ and Python. (2nd ed.)*. RA-MA Editorial.

[21]. Carranza, R. E. (2021). Business Intelligence and its influence on decision-making in the Academic area of IESTPAY - YURIMAGUAS. *Institutional Repository - UCV*. Retrieved from: https://repositorio.ucv.edu.pe/handle/20.500.12692/75346 [accessed: 06 February 2024].

[22]. Capuñay, D. L. (2021). Neural network-based model to project the academic performance of second grade of secondary school at Educational Institution No. 16093-Jaén. *Institutional Repository - UCV*. Retrieved from: https://repositorio.ucv.edu.pe/handle/20.500.12692/69442 [accessed: 06 February 2023].