Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.10 n.49, p. 222-227. September/October., 2024.
DOI: https://doi.org/10.5935/jetia.v10i49.1308

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

PREDICTING ACADEMIC PERFORMANCE IN UNIVERSITY STUDENTS USING MACHINE LEARNING

Ernesto Bolaños-Rodríguez*[1], Cristina Flores-Amador[2], Asdrúbal López-Chau[3], Alonso Ernesto Solis-Galindo[4] and Antonio Zárate-Rosas[5]

[1,2,4,5] Escuela Superior de Tizayuca-Universidad Autónoma del Estado de Hidalgo. Carretera Federal Tizayuca-Pachuca Km 2.5. 43800. Tizayuca, Estado de Hidalgo, México.
[3]Centro Universitario UAEM Zumpango, Universidad Autónoma del Estado de México, 55600, México.

[1] http://orcid.org/0000-0002-1432-7720 , [2] http://orcid.org/0000-0001-8122-3094 , [3] http://orcid.org/0000-0001-5254-093
[4]http://orcid.org/0000-0002-3999-006X , [5] http://orcid.org/0009-0005-9851-0992

Email: ebolanos@uaeh.edu.mx, floresa@uaeh.edu.mx, alchau@uaemex.mx, soliser@uaeh.edu.mx, za447459@uaeh.edu.mx

ARTICLE INFO

Article History
Received: September 26th, 2024
Revised: October 21th, 2024
Accepted: October 22th, 2024
Published: October 31th, 2024

Keywords:
Academic Performance,
Prediction,
Supervised Algorithms,
Machine Learning.

ABSTRACT

In the present research, the prediction of the academic performance of university students of an undergraduate educational program is carried out by applying Machine Learning (ML) with the purpose of determining the students with academic difficulties and excellence in school performance. It is an applied research in a population of 327 students, to which a representative sample of 74 students is determined, using a proportional stratified probability sampling, in which the stratum is the semester the student is studying out of the nine that make up the study plan. The work is an applied study with a pre-experimental design of a single group, because after applying ML the results are observed and the measurements is carried out. The main conclusions obtained allow establishing a methodology for the application of ML methods in the prediction of academic performance. The best performing algorithms used are Support Vector Machine (SVM) and Neural Network (NN).

I. INTRODUCTION

Academic performance is multifactorial in nature [1],[2], because it involves factors endogenous and exogenous to the university student that influence it. Several investigations have been carried out, using inferential statistics, data mining and ML to study the dependent variable academic performance and its relationship with multiple independent variables linked to socioeconomic, cognitive and emotional factors, which have shown that their dependencies are a function of the characteristics of the internal and external context where the student is located, and their implications can vary significantly [3],[4].

Therefore, the objective of this research conducted in a public university for an educational program in economic and administrative sciences is to predict the academic performance of university students through the application of supervised ML algorithms to detect students with school difficulties and those of excellence and thus have an impact on the educational process. For which the following research questions are formulated and answered in this paper: Which supervised learning methods are most suitable for predicting academic performance, considering survey responses from students as input? How effective is the performance of the best supervised learning method for making this prediction? What kind of data processing is necessary to accurately predict student performance?

The main contributions of the research lie in establishing a methodology for the application of ML methods in the prediction of academic performance, as well as evaluating the performance of supervised algorithms used for this purpose through the metrics of precision, recall and F1-score, which confer reliability and validity to the results obtained.

II. THEORETICAL REFERENCE

II.1 ACADEMIC PERFORMANCE AND IMPLICATIONS

When we talk about academic performance, particularly with undergraduate students, it is oriented to the level of achievement that the student achieves from the criteria established from the school institutions themselves, where it seeks to address the learning and knowledge from an accurate perspective involving students in various teaching environments, in which different internal and external factors that intervene in the results of the abstract and concrete formative processes are also combined, which can limit or contribute to individual school performance, such as the personality of each subject, the scenarios in which he/she develops, and, mainly, that he/she acquires the appropriate knowledge, competencies and skills that allow him/her to develop in a competitive environment.

Thus, academic performance can be understood as the conjugation of different multicausal factors that affect academic results, in which sociodemographic, psychosocial, pedagogical, institutional and socioeconomic elements interact; among them, elements as varied as: motivation, anxiety, self-esteem, perception of the academic climate, enthusiasm, the teacher, sense of purpose and others [5].

In this sense, there are several components that can be considered in the academic results, which are not always the same, which is why it is an issue that requires full attention to provide an appropriate approach in the training of students based on the dominant knowledge acquired during their academic career, as well as the know-how and know-how to be, coupled with the university values received inside and outside the classroom, accompanied by mastery of appropriate tools, appropriate study habits, an assertive pedagogy and quality in the teaching received by the student body, This can affect the formative development of the student from the emotional aspect, impacting also on the institutional goals of graduation, desertion, backwardness, and the image projected to the outside, since the universities have the purpose of contributing to the insertion of high impact professionals in society, without affecting its positioning as an alternative educational offer.

Thus, academic performance is understood as a student's level of knowledge measured in an evaluation test. In addition to intellectual level, academic performance is influenced by personality variables (extraversion, introversion, anxiety) and motivational variables, whose relationship with academic performance is not always linear, but is modulated by factors such as level of schooling, sex and aptitude. Other variables that influence academic performance are interests, study habits, teacher-student relationship, and self-esteem [6].

In terms of academic performance in higher education, the factors associated with its valuation and the results obtained by university students are based on the sum of quantitative elements that are related to what has been learned and what has been demonstrated to have been learned, whose results derive from a sum of different aspects and activities that are conditioned to the student body for a determined period of time, argued in a systematized context, contextualized in methodologies with pertinent approaches for the achievement of successful results.

In this sense, it can be mentioned that academic performance, being multicausal, involves an enormous explanatory capacity of the different factors and temporal spaces that intervene in the learning process, hence, there are different aspects associated with academic performance, among which both internal and external components of the individual intervene. They can be social, cognitive and emotional, which are classified into three categories: personal determinants, social determinants and institutional determinants, which present subcategories or indicators [7].

The importance for universities of measuring the academic performance of the school community represents the fulfillment of their educational task, commitments and mission to society, and also allows them to evaluate the quality standards achieved as a public or private educational institution, which largely determines the social image it has. In this way, the results achieved by students progressively during and at the end of their professional training ratifies them to identify with greater precision the knowledge, skills, abilities and knowledge that they have actually received to develop in the labor field, affirming the strengths they have and areas of opportunity that should be improved, which contributes to decision-making from the methodologies applied in the teaching-learning process and in the internal and external factors that influence school performance, which will affect the context in which the graduates are inserted into the labor market.

II.2 MACHINE LEARNING

Machine Learning (ML) is a field of artificial intelligence focused on developing, designing, adapting, or improving algorithms that enable computers to autonomously solve specific tasks. The key tasks in ML include classification, regression, clustering, anomaly detection, and association rules. For this article, supervised algorithms were employed to predict academic performance.

The specific algorithms used in the research are: Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Trees (DT) and Neural Network (NN), mainly because they are the most used in the research reported for this purpose [8] and they are the ones that present more possibilities and better performances to predict the response variable (academic performance) from a large number of input data. The main characteristics of each of them are described below in a synthesized form [9],[10]:

Logistic Regression (LR)

It is a type of classification analysis used to predict the outcome of a categorical variable as a function of independent variables.

Random Forest (RF)

It is a data algorithm that can be used in regression or classification tasks, it helps data science to make decisions from a series of questions that allow the final decision.

Support Vector Machine (SVM)

Support Vector Machines is a supervised learning algorithm that is based on finding a hyperplane that best separates different classes of data points. We use a trained SVM regression model due to the low dimensional predictor dataset.

K-Nearest Neighbors (KNN)

It is a nonparametric supervised learning classifier that uses proximity to make classifications and predictions about the clustering of an individual data point.

Decision Trees (DT)

Algorithm used in predictive modeling, which makes forecasts based on the relationships established between input and output data for decision making.

Neural Network (NN)

Neural network models are structured as a series of layers to mimic the way the brain processes information. We use a fully connected feedforward NN. The first layer of the NN has a connection from the network input (predictor data), and each subsequent layer has a connection from the previous layer. The final fully connected layer produces the output of the network, that is, the predicted response values.

Research on the use of ML in the prediction of academic performance has been reported, among which are those conducted by Chauhan and collaborators (2019) in which a machine learning tool is created to predict the grade point average based on past data. Theory and practice grades are taken, then regression techniques such as KNN, DT, SVM, RF and Linear Regression are applied. As a result, the relevance of comparing techniques to obtain a model with better accuracy, in this case multiple linear regression, is achieved [11].

On the other hand, Canagareddy and collaborators (2018) have proposed predictive models that allow predicting student performance so that corrective actions can be taken. Classification algorithms (NaireBayes, Logistic Classifier and T48 Classifier) and prediction algorithms such as SVM, RF and LR are combined in this study, resulting in this research the algorithm with the highest effectiveness the RF [12].

Also, Burman (2019) investigates how to help students improve their academic performance with the use of applications based on data mining, where a predictive model is created with SVM that allows classifying students into three categories: High, Medium and Low [13].

Finally, Candia (2019) conducts a prediction study of student academic performance from entry data using ML. For the development of the predictive model, the CRISP-DM methodology, the weka tool and demographic and educational factors are considered. RF, KNN and NN algorithms are employed, with RF achieving the best performance [14].

III. MATERIALS AND METHODS

III.1 POPULATION, SAMPLE AND SAMPLING

It is an applied research with a study population of 327 students enrolled in the July-December 2024 school year of the Bachelor's Degree in Tourism from first to ninth semester.

Within this approach, the following statistical mathematical formula is applied to determine the representative sample size of the population. See equation 1, [15]:

$$n= \frac{Z_{\alpha}^{2} \cdot N \cdot p \cdot q}{i^{2}(N-1) + Z_{\alpha}^{2} \cdot p \cdot q} \tag{1}$$

Where:
n= Sample
N= Size of the population or universe. Having 327 students.
i= Unforeseen or uncontrolled error, takes values from 1 to 10%
p= Probability of success (0.5)
q= Probability of failure (0.5)
Z= Gaussian Normal Curve statistic, at 95% confidence and 5% error. Z = 1.96 [15].

The sample consisted of 74 students, to whom the data collection instrument (designed questionnaire) was applied.

It is using a proportional stratified probability sampling, in which the stratum is the semester the student is studying out of the nine that make up the study plan.

III.2 DATA COLLECTION INSTRUMENT

A 28-question questionnaire is used, considering the following factors: personal, self-concept, motivational, sociocultural, parental education, emotional intelligence, economic, school of origin and average academic performance, which gives validity to the data collection instrument because it measures what it must measure [2-6].

The reliability of the questionnaire was determined by calculating Cronbach's Alpha coefficient with the support of SPSS for Windows Version 26.0.

III.3 METHODOLOGY USED

An Exploratory Data Analysis (EDA) was conducted to uncover potential complexities that could impact the performance of predictive models. Subsequently, the data was processed to address these identified issues. The processing steps included handling missing values, selecting the most relevant features for prediction, and balancing class distributions using the SMOTE algorithm. Additionally, numerical features were standardized by removing the mean and scaling to unit variance, while categorical features were encoded using One-Hot Encoding (OHE), except for the target variable. Python programming language, scikit and imblearn libraries were used.

The target variable is numeric; therefore, it was transformed into a categorical variable using the traditional grading scale employed in Mexican public schools. Consequently, the numeric values were replaced with the following labels: 'Fail', 'Poor', 'Fair', 'Good', 'Very Good', and 'Excellent'. See Table 1.

Table 1: Scale used to categorize the range of average academic performance data.

Score	Category
$0 \leq$ score < 6	'Fail'
$6 <$ score ≤ 7	'Poor'
$7 <$ score ≤ 8	'Fair'
$8 <$ score ≤ 9	'Good'
$9 <$ score ≤ 9.5	'Very Good'
$9.5 <$ score ≤ 10	'Excellent'

Source: Authors, (2024).

III.4 MACHINE LEARNING FRAMEWORK

Figure 1 illustrates the applied framework. In the methodology, the collected data is split into two sets: a training set and a testing set. The testing set comprises 20% of the data, while the remaining 80% is used for training. The data undergo preprocessing, which involves selecting the most important features, standardizing numerical variables, and applying One-Hot Encoding (OHE) to categorical variables. Additionally, class balancing is performed. The processed data is then used to train classification models. During model training, the optimal parameters for each model are selected using grid search and cross-validation techniques. The best-performing models of each classifier type are then evaluated on the testing set. The testing data is processed using the same techniques as the training set, except for class balancing.

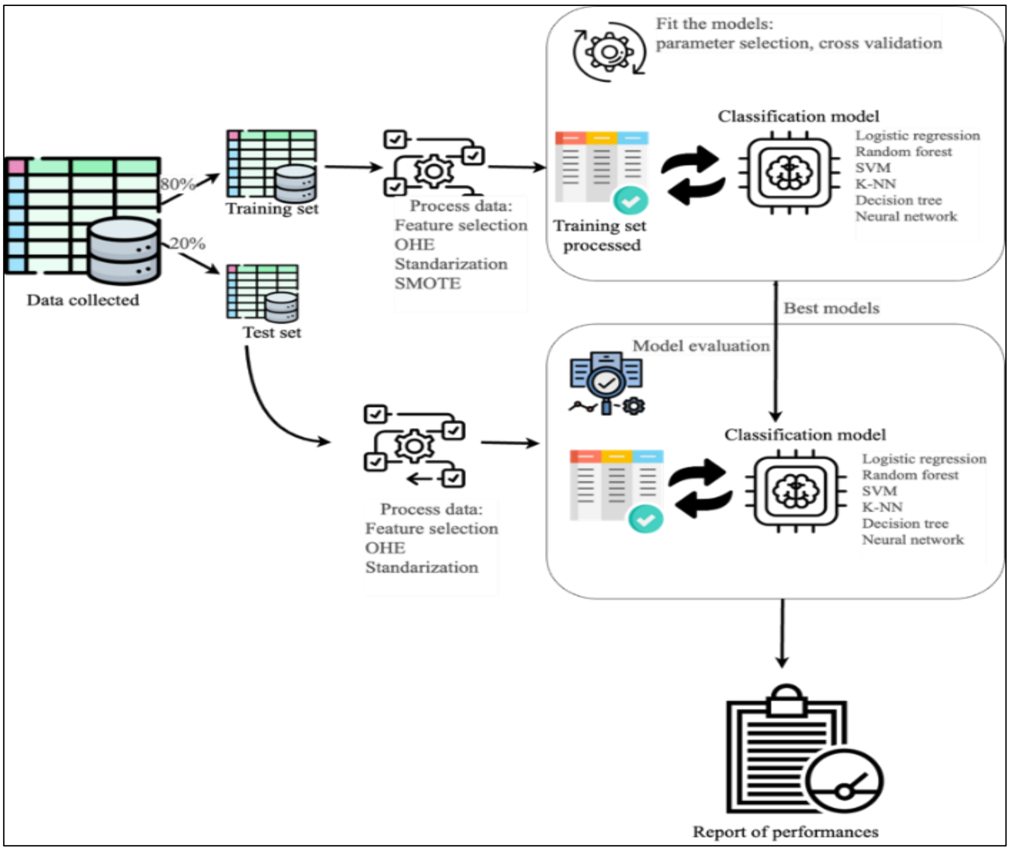

Figure 1: Methodology used for the application of machine learning methods in the prediction of academic performance.
Source: Authors, (2024).

Each model is assessed using various metrics, including precision, recall, F1-score, and accuracy. The results are reported in a later section. Accuracy, which consists of obtaining the proportion of correct predictions, both true positives and false positives, over the total number of predictions. This metric is mainly used to measure the overall performance of the whole model. The sensitivity metric, which consists of finding the proportion of true positives over the total number of true positives, is also used to measure the model's ability to correctly identify positive instances. Finally, the F1-score, which is the harmonic mean between accuracy and recall, is used to find which class best predicts the model [16].

IV. RESULTS AND DISCUSSIONS

The results obtained for Cronbach's Alpha Coefficient is 0.93, which means that the questionnaire applied is reliable, i.e. it denotes stability, consistency, minimum errors, in the data collection instrument; if the questionnaire were repeated under the same conditions, the results would be similar. As a result of the exploratory data analysis (EDA), 59 records with complete data were identified and retained, while records with missing data were removed. A significant data imbalance was observed, with "Excellent" being the minority class. Figure 2 illustrates the proportion of each category. This finding is not surprising, as students with excellent academic performance are relatively few. The "Good" performance category is the most prevalent in the dataset. No instances of the "Fail" or "Poor" classes were found.

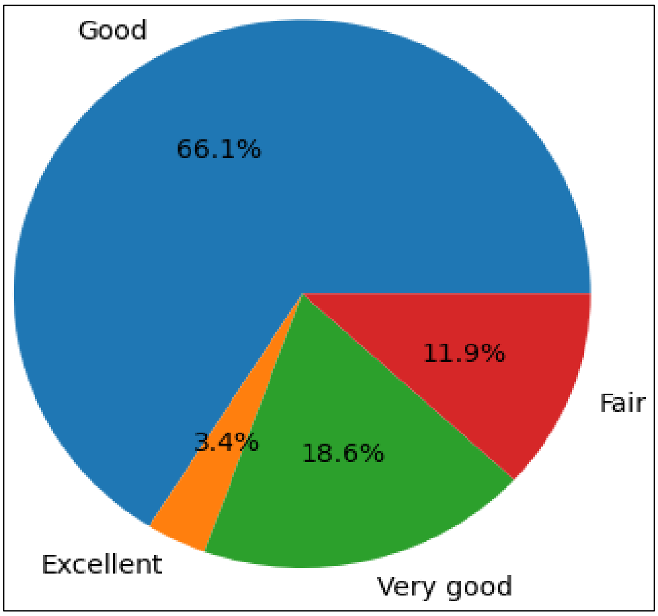

Figure 2: Distribution of academic performance categories in the dataset.
Source: Authors, (2024).

The imbalance among the categories led to poor performance in the classification methods, necessitating the balancing of classes within the training dataset. Following the application of the SMOTE algorithm to the training data, the class distribution shown in Figure 3 was achieved, which demonstrates a balanced representation across all categories.

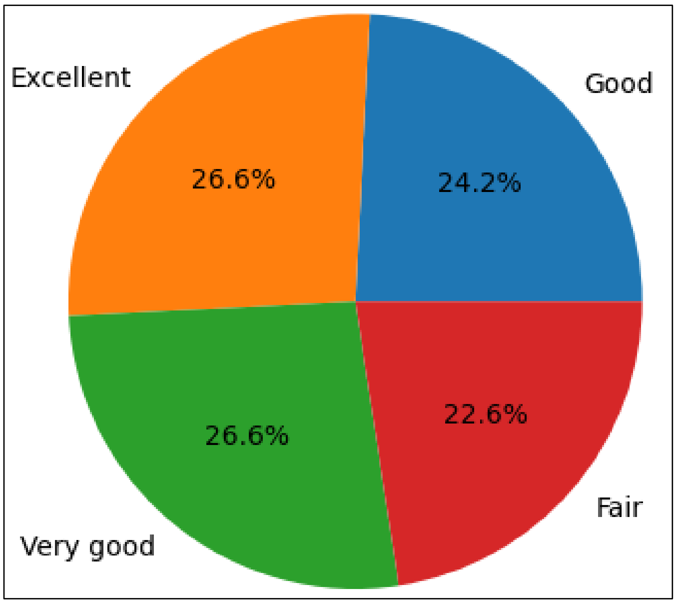

Figure 3: Distribution of categories in the training dataset after applying SMOTE.
Source: Authors, (2024).

Applying the feature selection method, the following attributes were identified as relevant: age, age at the start of the program, semester, weekly study hours, weekly working hours, family members, and monthly income. Other variables were excluded. All categorical variables were encoded using One-Hot Encoding (OHE), while numerical variables were standardized. This process resulted in a dataset with 70 attributes.

The results of the six evaluated classifiers are presented below in Tables 2-7. The models that achieved perfect prediction (100%) are the Support Vector Machine (SVM) and the Neural Network. The second-best classifier is Random Forest, with an accuracy of 88%. K-Nearest Neighbors and Decision Tree classifiers were the models with the lowest performances.

Table 2: Performance of Logistic Regression classifier with optimal parameter C=1.0.

	Precision	**Recall**	**F1-score**
Excellent	1.00	1.00	1.00
Fair	0.79	1	0.88
Good	1.00	0.44	0.62
Very good	0.75	1.00	0.86
Accuracy	-	-	0.84
macro avg	0.88	0.86	0.84
weighted avg	0.88	0.84	0.82

Source: Authors, (2024).

Table 3: Performance of Random Forest classifier with parameters n_estimators=50.

	Precision	**Recall**	**F1-score**
Excellent	1.00	0.83	0.91
Fair	1.00	1.00	1.00
Good	0.78	0.78	0.78
Very good	0.71	0.83	0.77
accuracy	-	-	0.88
macro avg	0.87	0.86	0.86
weighted avg	0.88	0.88	0.88

Source: Authors, (2024).

Table 4: Performance of Support Vector Machine classifier with parameters 'C'=1, kernel=rbf, gamma=0.1.

	Precision	**Recall**	**F1-score**
Excellent	1.00	1.00	1.00
Fair	1.00	1.00	1.00
Good	1.00	1.00	1.00
Very good	1.00	1.00	1.00
accuracy	-	-	1.00
macro avg	1.00	1.00	1.00
weighted avg	1.00	1.00	1.00

Source: Authors, (2024).

Table 5: Performance of K-Nearest Neighbors classifier with parameters n_neighbors=3, weights=distance.

	Precision	**Recall**	**F1-score**
Excellent	0.86	1.00	0.92
Fair	0.85	1.00	0.92
Good	1.00	0.11	0.20
Very good	0.55	1.00	0.71
accuracy	-	-	0.75
macro avg	0.81	0.78	0.69
weighted avg	0.84	0.75	0.68

Source: Authors, (2024).

Table 6: Performance of Decision Tree classifier with parameters max_depth=10, min_samples_split=10.

	Precision	**Recall**	**F1-score**
Excellent	1	0.83	0.91
Fair	0.91	0.91	0.91
Good	0.67	0.67	0.67
Very good	0.57	0.67	0.62
Accuracy	-	-	0.78
macro avg	0.79	0.77	0.78
weighted avg	0.79	0.78	0.79

Source: Authors, (2024).

Table 7: Performance of Neural Network classifier with parameters activation=ReLu, hidden_layer_sizes= (100, 100, 100), max_iter=500, solver=adam.

	Precision	**Recall**	**F1-score**
Excellent	**1.00**	**1.00**	**1.00**
Fair	1.00	1.00	1.00
Good	1.00	1.00	1.00
Very good	1.00	1.00	1.00
accuracy	-	-	1.00
macro avg	1.00	1.00	1.00
weighted avg	1.00	1.00	1.00

Source: Authors, (2024).

In this way, the questions posed at the beginning of this article are answered below.

Which supervised learning methods are most suitable for predicting academic performance, considering survey responses from students as input? The most suitable methods are SVM and NN.

How effective is the performance of the best supervised learning method for making this prediction? The two best supervised learning methods are SVM and NN, achieving perfect performance according to the metrics used to measure their effectiveness.

What kind of data processing is necessary to accurately predict student performance? It is necessary to conduct an exploratory data analysis (EDA) to identify potential complexities, and it is also advisable to address them properly. Additionally, choosing an appropriate encoding for categorical variables is important.

V. CONCLUSIONS

A methodology is established for the application of ML methods in the prediction of academic performance. It is possible to predict the academic performance of university students with the application of the supervised algorithms of Support Vector Machine (SVM) and the Neural Network, which have the best performances with respect to metrics precision, recall, F1-score and accuracy, followed by Random Forest, with an accuracy of 88%, while. K-Nearest Neighbors and Decision Tree classifiers were the models with the lowest performances. This will allow us to detect students with school difficulties in order to carry out actions that will have an effective impact on the educational process through strategies such as: academic counseling, tutoring, psychological care and financial support through scholarships, on the other hand, by predicting the students of excellence we will be able to follow them so that these students can participate in professional competitions and perform national and international mobility to Higher Education Institutions of prestige and recognition that will give visibility to our undergraduate educational program under study.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Ernesto Bolaños-Rodríguez and Asdrúbal López-Chau.
Methodology: Asdrúbal López-Chau, Alonso Ernesto Solis-Galindo and Antonio Zárate-Rosas.
Investigation: Ernesto Bolaños-Rodríguez, Cristina Flores-Amador, Asdrúbal López-Chau, Alonso Ernesto Solis-Galindo and Antonio Zárate-Rosas.
Discussion of results: Ernesto Bolaños-Rodríguez and Asdrúbal López-Chau.
Writing – Original Draft: Ernesto Bolaños-Rodríguez.
Writing – Review and Editing: Ernesto Bolaños-Rodríguez and Cristina Flores-Amador.
Resources: Cristina Flores-Amador.
Supervision: Cristina Flores-Amador and Asdrúbal López-Chau.
Approval of the final text: Ernesto Bolaños-Rodríguez, Cristina Flores-Amador, Asdrúbal López-Chau, Alonso Ernesto Solis-Galindo and Antonio Zárate-Rosas.

VII. ACKNOWLEDGMENTS

We would like to thank the directors, professors and students who collaborated in the development of this research at the Universidad Autónoma del Estado de Hidalgo and the Universidad Autónoma del Estado de México.

VIII. REFERENCES

[1] A. Chilca, 2017, Self-esteem, study habits and academic performance in college students. [Online]. Available at: https://revistas.usil.edu.pe/index.php/index/oai?verb=ListRecords&metadataPrefix=oai_dc&set=pyr:ART+INV

[2] G.M. Garbanzo Vargas, "Factors associated with academic performance in university students, a reflection on the quality of public higher education", Education Journal, vol. 31, no. 1, p. 43, 2012. https://www.scielo.sa.cr/scielo.php?pid=S1409-42582013000300004&script=sci_abstract

[3] C.E. López Guarin, E.L. Guzmán & F.A. González, "A Model to Predict Low Academic Performance at a Specific Enrollment Using Data Mining", Iberoamerican Journal of Learning Technologies, vol. 10, no. 3, pp. 119-125, 2015. doi: 10.1109/RITA.2015.2452632

[4] A. Mueen, B. Zafar & U. Manzoor, "Modeling and Predicting Students Academic Performance Using Data Mining Techniques. International", Journal of Modern Education and Computer Science, vol. 8, no. 11, pp. 36-42, 2016. doi: 10.5815/ijmecs.2016.11.05

[5] E. Montero Rojas, J. Villalobos Palma & A. Valverde Bermudez, "Institutional, pedagogical, psychosocial and sociodemographic factors associated with academic performance and student repetition at the University of Costa Rica", Institute of Psychological Research, University of Costa Rica, 2004. doi: https://doi.org/10.7203/relieve.13.2.4208

[6] F. Fajardo Bullón, M. Maestre Campos, E. Felipe Castaño, B. León del Barco, & M.I. Polo del Río, "Analysis of the academic performance of students in compulsory secondary education according to family variable", Education XX1, vol. 20, no. 1, pp. 209-232, 2017. doi: 10.5944/educXX1.14475

[7] S. Pelegrina, M.C. García & P. F. Casanova, "Parenting styles and adolescents' academic performance", Electronic Journal USAL, vol. 25, no. 2, pp. 147-168, 2002. doi: https://doi.org/10.1174/021037002317417796

[8] V.L. Uskov, J.P. Bakken, A. Byerly & A. Shah, "Machine Learning-based Predictive Analytics of Student Academic Performance in STEM Education", IEEE Global Engineering Education Conference (EDUCON) , Dubai, United Arab Emirates, 2019, pp. 1370-1376, 2019. doi: 10.1109/EDUCON.2019.8725237

[9] E. Bolaños Rodríguez, J.C. González Islas, O. López Ortega & E. Lezama León, "Application of Machine Learning for the Prediction of Atmospheric Corrosion in the Metropolitan Area of Mexico City", Journal of Engineering and Technology for Industrial Applications, vol. 10, no. 48, p. 36-42, July-August, 2024. doi: https://doi.org/10.5935/jetia.v10i48.1089

[10] L. Breiman, "Random Forests", Machine Learning, vol. 45, pp. 5–32, 2001. https://doi.org/10.1023/A:1010933404324

[11] N. Chauhan, K. Shah, D. Karn & J. Dalal, "Prediction of Students Performance Using Machine Learning", SSRN Electronic Journal, 2019. doi: http://dx.doi.org/10.2139/ssrn.3370802

[12] D. Canagareddy, K. Subarayadu & V. Hurbungs, 2018, A Machine Learning Model to Predict the Performance of University Students [Online]. S.I. Springer International Publishing. https://doi.org/10.1007/978-3-030-18240-3_29

[13] I. Burman & S. Som, "Predicting Students Academic Performance Using Support Vector Machine", Proceedings-2019 Amity International Conference on Artificial Intelligence, pp. 756-759, AICAI 2019. doi: https://doi.org/10.1109/AICAI.2019.8701260

[14] D. I. Candia Oviedo, 2019, Predicting the Academic Performance of UNSAAC students from their entrance data using machine learning algorithm, National University of San Antonio Abad del Cusco [Online] pp.141. Available at: https://repositorio.unsaac/4120

[15] J.T. McClave, P.G. Benson & T. Sincich, *Statistics for Business and Economics*, Pearson, Florida, 2022. https://api.pageplace.de/preview/DT0400.9781292413525_A42820209/preview-9781292413525_A42820209.pdf

[16] R. Borja Robalino, A. Monleon Getino & J. Rodellar, "Standardization of performance metrics for Machine and Deep Learning classifiers", Iberian Journal of Information Systems and Technologies, vol. 30, pp. 184-196, 2020. https://www.researchgate.net/publication/339943922