

Machine Learning Model through Ensemble Bagged Trees in Predictive Analysis of University Teaching Performance

Omar Chamorro-Atalaya¹
Facultad de Ingeniería y Gestión
Universidad Nacional Tecnológica
de Lima Sur
Lima- Perú

Marco Anton-De los Santos³
Juan Anton-De los Santos⁴
Facultad de Ciencias Económicas
Universidad Nacional Federico
Villarreal, Lima-Perú

Antenor Leva-Apaza⁶
Facultad de Ciencias
Universidad Tecnológica del Perú
Lima-Perú

Carlos Chávez-Herrera²
Facultad de Ingeniería de Sistemas e
Informática Universidad Nacional
Mayor de San Marcos
Lima-Perú

Almintor Torres-Quiroz⁵
Facultad de Ciencias Económicas
Universidad Nacional del Callao
Lima-Perú

Abel Tasayco-Jala⁷
Gutember Peralta-Eugenio⁸
Facultad de Ciencias Empresariales
and Facultad de Ciencias de la Salud
Universidad César Vallejo
Lima-Perú

Abstract—The objective of this study is to analyze and discuss the metrics of the Machine Learning model through the Ensemble Bagged Trees algorithm, which will be applied to data on satisfaction with teaching performance in the virtual environment. Initially the classification analysis through the Matlab R2021a software, identified an Accuracy of 81.3%, for the Ensemble Bagged Trees algorithm. When performing the validation of the collected data, and proceeding with the obtaining of the predictive model, for the 4 classes (satisfaction levels), total precision values of 82.21%, Sensitivity of 73.40%, Specificity of 91.02% and of 90.63% Accuracy. In turn, the highest level of the area under the curve (AUC) by means of the Receiver operating characteristic (ROC) is 0.93, thus considering a sensitivity of the predictive model of 93%. The validation of these results will allow the directors of the higher institution to have a database, to be used in the process of improving the quality of the educational service in relation to teaching performance.

Keywords—Machine learning; ensemble; bagged trees; predictive analysis; teaching performance

I. INTRODUCTION

The information and communication technology (ICT) sector is currently a leader in the analysis of data from different media [1], [2], such as virtual platforms, survey administration software, among other technological tools [3], [4], which capture or acquire information to be processed and analyzed in descriptive statistical research or in research on predictive models applicable to various areas of knowledge [5].

The advantages that the introduction of ICT has generated in the education sector is based on the importance of technology to develop research that previously could not be carried out, [6], [7] as is the case of the identification of predictive models for the analysis or monitoring of university

teaching performance, student performance, among other relevant factors for the education sector [8]-[10].

Worldwide, the education sector has undergone changes and transformations, due to the virtualization of the teaching-learning mode, [11], [12], [13], as a consequence of this scenario, universities face new challenges, to safeguard the quality of education that goes hand in hand with the advancement of technology [14]-[16].

Given this, in the education sector, an increasing amount of data has been generated with greater relevance, product of the iterations of the different actors of the educational process, these being the teacher, the students and the institution, through the application of tools technological, such as survey software, which generate a database [17], [18]. As indicated, the data that are stored, are used in order to improve the efficiency of the educational process through predictive models, among the factors to optimize are academic performance, student dropout, teaching performance, graduate follow-up [19].

There are various technologies used to obtain predictive models, which use data from virtual platforms and survey administration software, applied to students by universities [20]. Within these technologies is the branch of Artificial Intelligence that within its fields houses Machine Learning [21]-[23]. As indicated in [24], Machine Learning is a set of algorithms capable of learning to perform certain tasks from the generalization of examples. Machine Learning has been successfully applied to a variety of areas of human endeavor, and has recently been applied to the educational sector, whose purpose is oriented towards the design of algorithms, methods and models, which will allow the exploration of data from teaching-learning environments [25], [26].

Among the multiple algorithms of Machine Learning, there is Ensemble Bagged Trees, which is an algorithm that is used in joint learning [27]. This can combine training and base

classifiers to produce ensemble models or use an algorithm with multiple test data sets as the basis [28]. In this regard, in [29] it is pointed out that the Bagged Trees algorithm forms different trees when there is a change in the starting point of the training data that results in a decrease in stability. This technique or algorithm is also suitable to be used in the search for optimal models for large data, since the classification becomes easier [30], [31].

In this sense, the main objective of this article is to determine the predictive model using Machine Learning through the Ensemble Bagged Trees algorithm, for the predictive analysis of university teaching performance, in order to use it as part of the procedure to improve the quality of the educational process. Initially, the methodology used will be detailed, then the validation of the algorithm will be determined, by means of the accuracy and the confusion matrix, to finally analyze the total performance metrics (Accuracy (A), Precision (P), Sensitivity (S) and Specificity (R)) of the selected algorithm, from obtaining the receiver operating characteristic curve (ROC).

The contribution of the research focuses on applying a novel technique for the higher institution, through machine learning making use of the data and information collected, which allows making preventive and corrective decisions based on reliable results, obtained through a methodology not so complex.

II. RESEARCH METHODOLOGY

A. Type and Level of Research

The type of research is applied, since it starts from the identification of a problem, related to the improvement of university teaching performance, for which use is made of methods or tools already defined such as predictive models through Machine Learning, which employs the Ensemble Bagged Trees algorithm. Likewise, the research level is descriptive, since it focuses on analyzing and discussing the metrics of the predictive model obtained through the Ensemble Bagged Trees algorithm, applied to the perception data of engineering university students.

This research also seeks to design a predictive multidimensional model that can be used to create and store new data for the higher institution. Based on this technological tool, it determines patterns and calculates association rules, providing support and reliability to the results obtained. Performance metrics such as Accuracy, Precision, Sensitivity and Specificity show improved performance over the manual method of the same procedure commonly performed in research [28].

B. Participants

The participants in this research are made up of students from the sixth to the tenth cycle of professional engineering schools, with a total of 581 students, this selection criterion is part of a regulation established and approved by the higher institution. It should be noted that it was possible to collect data from the entire population, for this reason, it can be noted that the sample coincides with the population.

C. Data Collection Technique and Instrument

The data collection technique is the survey, and the instrument used to collect data regarding university teaching performance is the questionnaire, which was carried out virtually, due to the context of the health emergency declared by the Covid -19. The virtual platform of the higher institution was used, which gave access to the data collection instrument through the code of each student, which guaranteed the security and reliability of the information. The questionnaire consisted of responses on a Likert scale ranging in levels from 1 to 4 (from dissatisfied to very satisfied). These levels of satisfaction in the analysis will be represented as the classes of the predictive model. In Fig. 1, the indicators considered as predictive elements in the perception of university teaching performance are shown.

D. Reliability of the Collected Data

As part of the methodology, the validation of the collected data is carried out, through Cronbach's Alpha coefficient using the SPSS software, this analysis carried out, it is observed in Table I that the consistency coefficient is equal to 0.932. As indicated in [12], values greater than 0.9 indicate great consistency that is, high homogeneity and equivalence of the response of all indicators. Once this result is obtained, the following section shows the results.

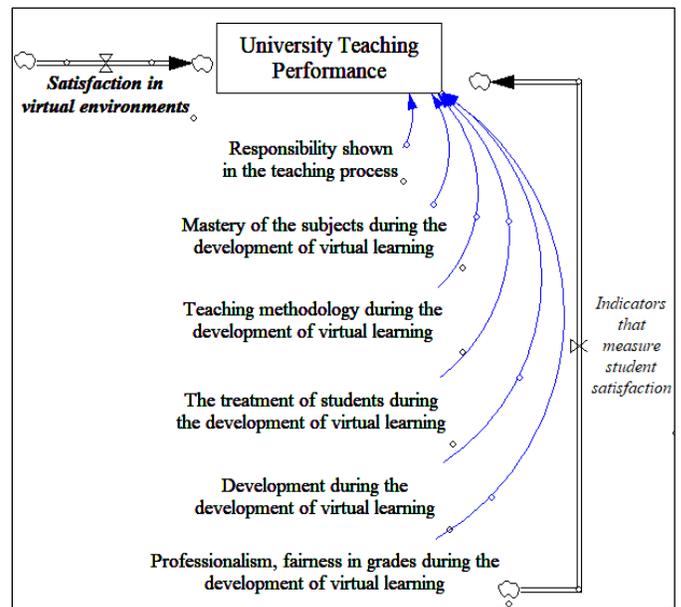


Fig. 1. Indicators that Measure Student Satisfaction with Teaching Performance.

TABLE I. CRONBACH'S ALPHA TEST

Reliability statistics	
Cronbach's alpha	No. of elements
0.932	6

E. Data Processing Design

The data processing design responded to a non-experimental transactional process, in which data was collected through a virtual questionnaire. In Fig. 2, the methodology of the research process is shown, which begins with the collection of data on the perception of engineering students from a public university in Peru. These data are related to the 6 indicators that are visualized in Fig. 1, whose appreciation regarding teaching performance is of an ordinal qualitative type, thus establishing 4 classes (very satisfied: 4, satisfied: 3, not very satisfied: 2 and dissatisfied :1).

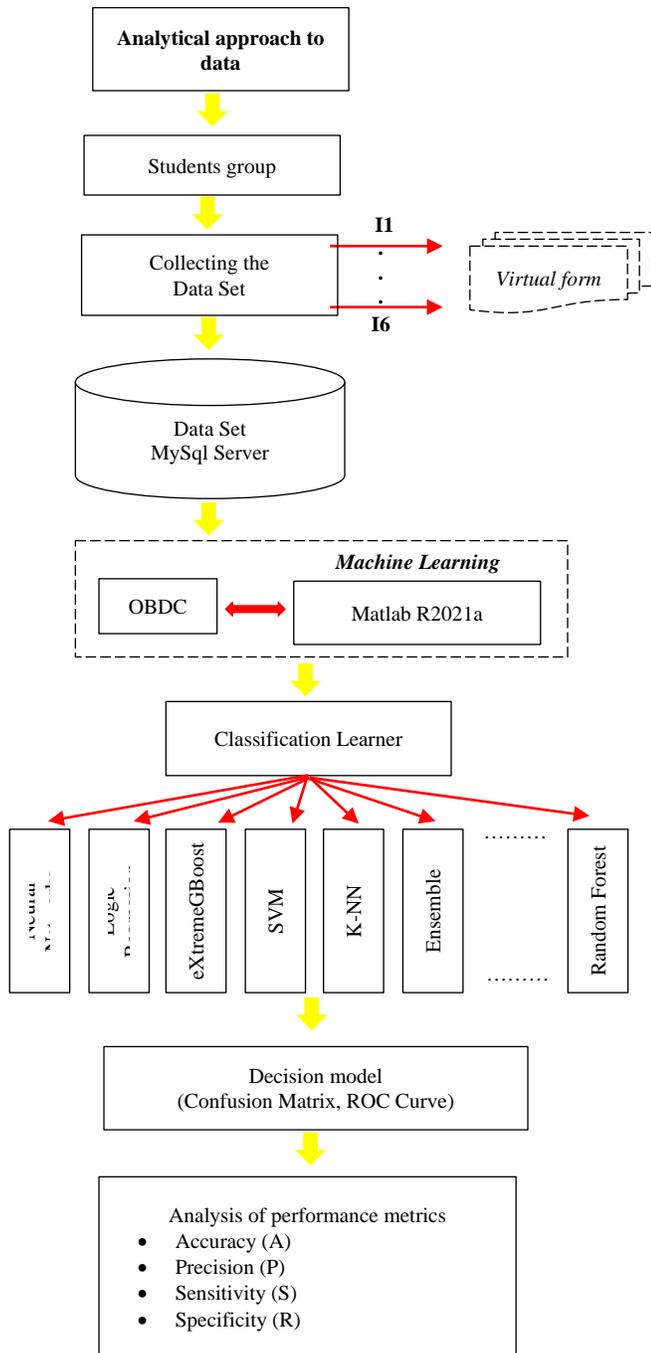


Fig. 2. Methodology of the Research Process through Machine Learning.

Likewise, the information collected was stored in a database in Microsoft SQL Server, associated through the Open Data Base Connectivity (ODBC) driver and the Matlab R2021a software. Using the Matlab software, we proceeded to use the “Classification Learner” tool, in order to identify the best Machine Learning algorithm, through its metrics. This algorithm allows the classification of students from the results obtained from the indicators specified in Fig. 1.

III. RESULTS AND DISCUSSION

A. Determination of the Predictive Model

Using the Matlab R2021a software, and using the Classification Learner and Statistics and Machine Learning Toolbox 12.1 application, the best predictive model determined by the validation of the accuracy is identified, in Fig. 3, the results generated by the software are shown. Matlab R2021a.

As shown in Fig. 3, the Machine Learning algorithm that presents the best accuracy, for classifying the level of satisfaction with respect to university teaching performance, is the Ensemble Bagged Trees algorithm with an accuracy of 81.3%.

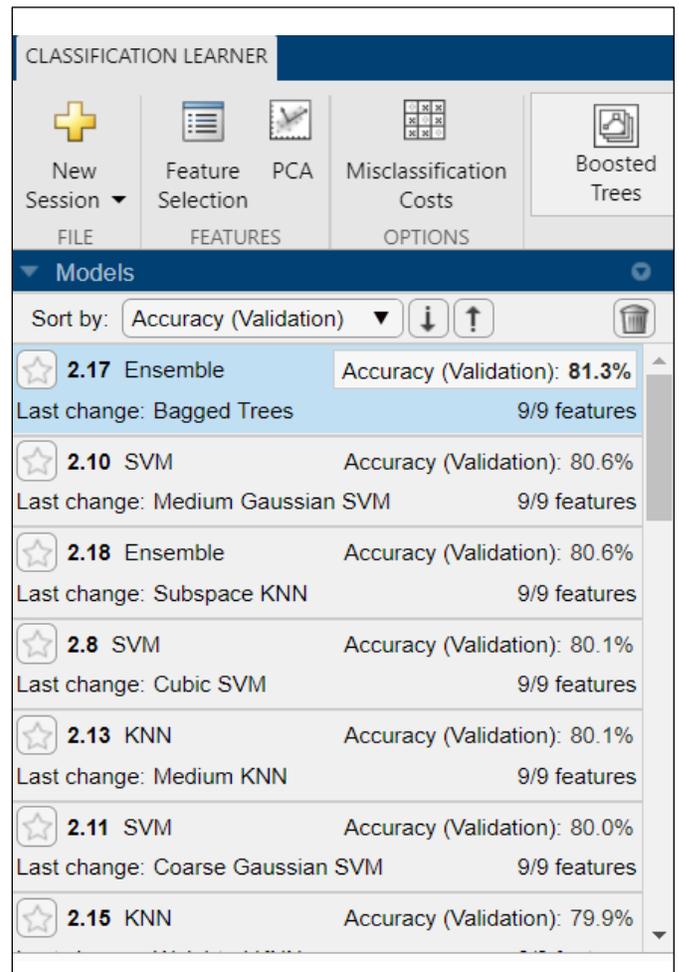


Fig. 3. Validation of the Prediction Algorithms Ordered by their Accuracy.

B. Results of the Predictive Model Metrics

When using the predictive model through Machine Learning through Ensemble Bagged Trees, to determine satisfaction with university teaching performance, confusion matrices are obtained, which represent elements of validation or performance measurement of the predictive model.

In Fig. 4, the confusion matrix is shown, with respect to the sensitivity metric, in it you can visualize the number of observations made by the classification system, and it reports the number of false negatives (FNR), which is the number of positive examples wrongly classified as negative and true positives (TPR) that define the number of positive samples correctly classified as positive, which shows the closeness between the levels of satisfaction predicted (Predicted class) by the model with respect to its true value (True class).

As can be seen in Fig. 4, of the 4 classes on which the predictive model acts through Ensemble Bagged Trees, class 3 shows the highest percentage of sensitivity, this means that the predictive model has the ability to discriminate between a true positive (TP) of a false negative (FN) in this class (satisfied), in this case it is 89.9%, as observed in this class the model was only confused by 10.1%. While the lowest level of sensitivity of the predictive model is shown in class 1 (satisfaction level: dissatisfied), whose value is 63.9%.

In Fig. 5, the confusion matrix is shown with respect to the precision metric, since the values of the main diagonal indicate the precision of the predictive model for each class.

In Fig. 5, the confusion matrix is shown regarding the precision metric of the predictive model for each class, in which it is observed that the predictive model for class 1 (satisfaction level: dissatisfied) shows the highest precision rate, in this case it is 88.5%. This result indicates that the level of dispersion of the data for this class is very low.

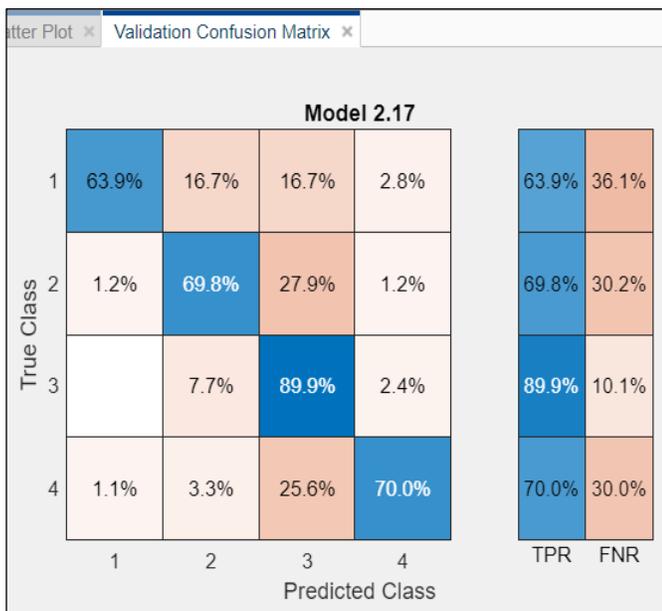


Fig. 4. Confusion Matrix based on TPR and FNR rates.

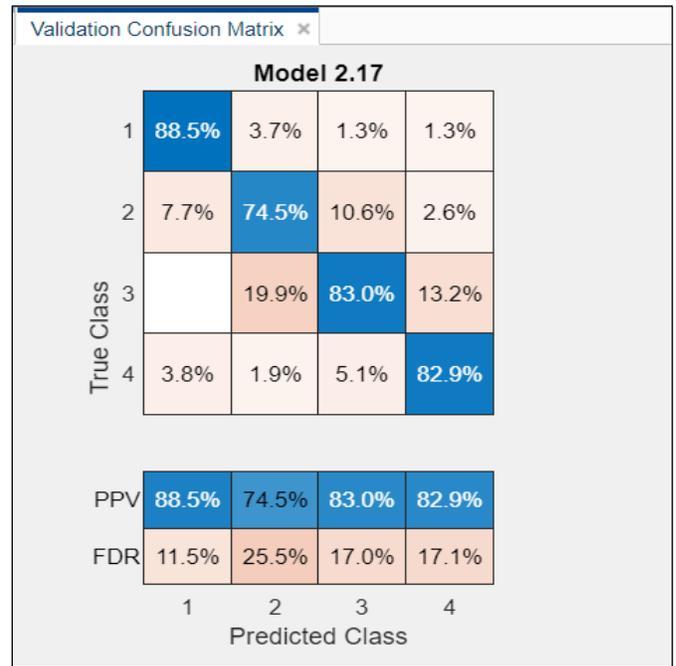


Fig. 5. Confusion Matrix based on PPV and FDR rates.

Table II shows the metrics of the predictive model through Ensemble Bagged Trees, for each class, in which it is evidenced that the total Precision is 82.21%, the total Sensitivity is 73.40% and the total Specificity is 91.02%, and the Accuracy presents a total value of 90.63%.

As part of the predictive model through the Ensemble Bagged Trees algorithm, the response that Matlab provides for each class under study is evidenced, its corresponding Receiver operating characteristic (ROC) graph and considering that the ROC graph describes the Sensitivity and Specificity of the algorithm classifier, the findings in Fig. 6, allow us to establish that for class 1 (dissatisfied), a sensitivity of 93% is shown.

In addition, the discrimination threshold is 0.64 for the rate of true positives and 0.00 for the rate of false positives, showing an area value on the curve (AUC) of 0.93, this value being close to 1, it is noted that the model for class 1 is optimal.

In Fig. 7, the ROC graph for class 2 (not very satisfied) is shown, where a sensitivity of 91% is displayed. In addition, the discrimination threshold is 0.70 for the rate of true positives and 0.08 for the rate of false positives, showing an area value on the curve (AUC) of 0.91, this value being close to 1, it is noted that the model for class 2 is optimal.

TABLE II. CLASSIFICATION PREDICTIVE ALGORITHM METRICS

Class	Metrics			
	Sensitivity	Specificity	Accuracy	Precision
1	63.89%	99.56%	97.76%	88.46%
2	69.77%	92.45%	86.99%	74.53%
3	89.93%	74.16%	83.36%	82.96%
4	70.00%	97.92%	94.41%	82.89%

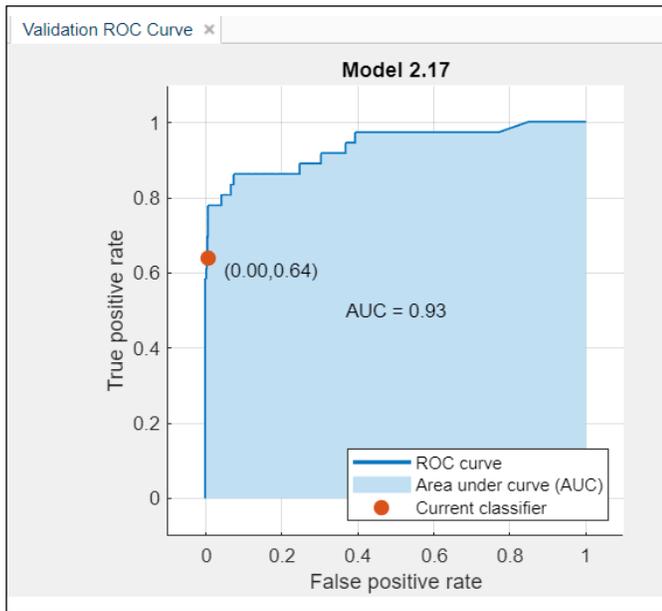


Fig. 6. ROC Charts for Class 1.

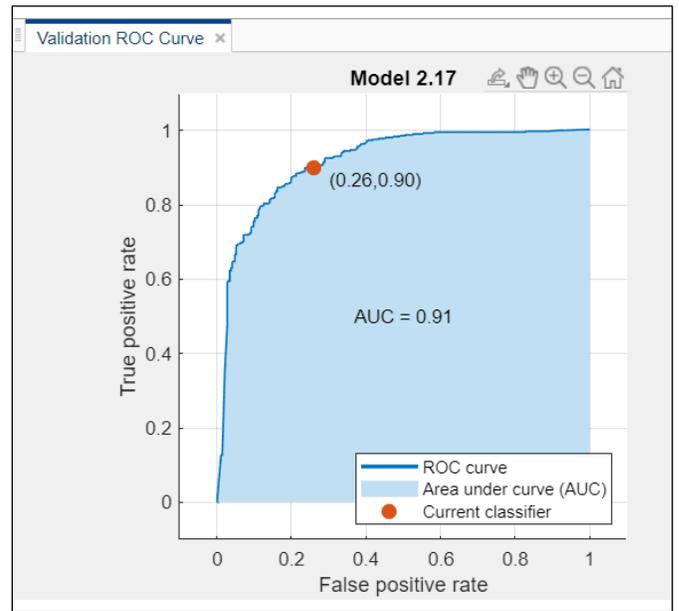


Fig. 8. ROC Charts for Class 3.

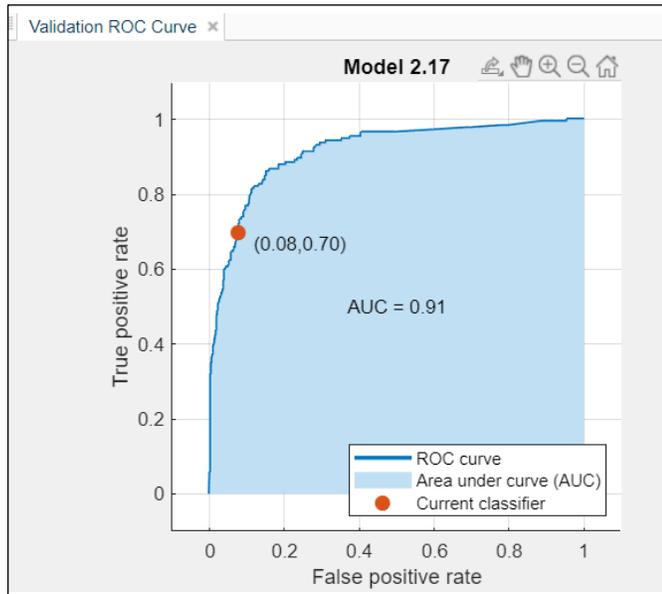


Fig. 7. ROC Charts for Class 2.

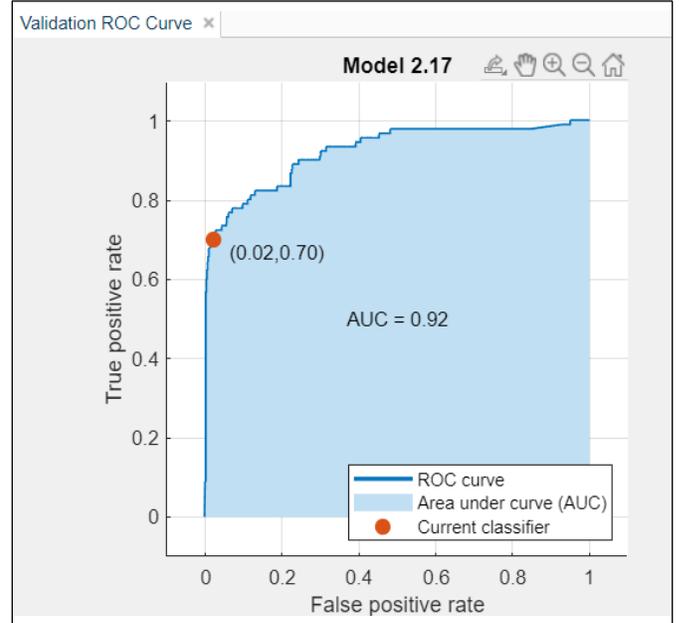


Fig. 9. ROC Charts for Class 4.

In Fig. 8, the ROC plot for class 3 (satisfied) is shown, where a sensitivity of 91% is displayed. In addition, the discrimination threshold is 0.90 for the rate of true positives and 0.26 for the rate of false positives, showing an area value on the curve (AUC) of 0.91, this value being close to 1, it is noted that the model for class 3 is optimal.

Finally, in Fig. 9, the ROC graph for class 4 (very satisfied) is shown, where a sensitivity of 92% is displayed. In addition, the discrimination threshold is 0.70 for the rate of true positives and 0.02 for the rate of false positives, showing an area value on the curve (AUC) of 0.92, this value being close to 1, it is noted that the model for class 4 is optimal.

C. Discussion

In relation to the results obtained, it is evidenced that the predictive model, based on the Ensemble Bagged Trees algorithm, presents acceptable metrics of precision, sensitivity, specificity and accuracy, in its 4 classes each of its classes, in this way the predictive model obtained provides security and reliability, contributing to decision making to improve the quality of the course content and the pedagogical methodology. In this regard, in [16] it is pointed out that preventive and corrective decision-making in higher education institutions involves building predictive models based on intelligent systems.

As indicated in [6], researchers have been concerned in recent years to work on the development of models that allow understanding aspects of the academic life of the student, teachers and institutions that allow the preparation and making of correct decisions, for the improvement continuity of educational quality. Likewise, in [19] it is indicated that the results obtained and validations show a precision of 82%, therefore, it can be pointed out that the process describes an optimal performance of the algorithms, so its incorporation would be satisfactory to be incorporated to the management of virtual educational knowledge.

In relation to the metrics of the predictive model, the model obtained through Matlab R2021a presents a general precision of 82.21% and an accuracy of 90.63%, being considered an optimal model, in this regard in [20], the author states that his predictive model was good since its general precision was 75.42% and an area under the ROC curve of 0.805. Likewise, in the investigation of [27] it is pointed out that the general result shows that each of the techniques used shows a good result in the classification and prediction performance, obtaining a greater precision of 86.9%.

On the other hand, the results of [26] showed a precision rate of 89.31% and a specificity rate of 91.25%, these measures are substantial to select classifiers since the researcher intends to minimize false negatives.

Regarding the term optimal model, in [4] it is pointed out that the so-called optimal models are combined with the dominant sets, which significantly improve the performance of prediction models and are highly influential in academic performance factors. Likewise, regarding the area on the curve, whose highest value in this research was 0.93 or 93%, in [4] it is indicated that an AUC of 50% of 91% or 99%, which was obtained in the research represents a better Classifier algorithm performance, favorable results for research.

The results of this study, from the perspective of innovation, will make it possible to achieve great changes, delegating functions, promoting competencies and fostering the continuous updating of higher institutions, all from the perspective of visionary leadership. In [10] it is pointed out that the proposed model accurately predicts the completion of the course and the performance of students in the university, thus allowing the organization to provide a better quality of service, since the satisfaction of the student depends on it student.

IV. CONCLUSION

The use of technological tools such as Machine Learning and its algorithms are supporting and strengthening decision-making from an administrative and academic point of view and in the educational sector. According to the results obtained, it is concluded that the metrics of the Machine Learning model through Ensemble Bagged Trees, applied to the predictive analysis of university teaching performance, present on average optimal values in their validation metrics such in their 4 classes, with a precision of 82.21%, a Sensitivity of 73.40%, a Specificity of 91.02% and an Accuracy of 90.63%. From the validation of the Machine Learning algorithm metrics, its implementation is viable and reliable in improving the performance of university teachers. Finding the 4 classes of the

predictive model with relatively high values, the results allow establishing the grouping of engineering students who can achieve a level of satisfaction based on the indicators called predictors (indicators), through which the authorities of the higher institution can make timely decisions to improve the percentage of satisfied students in relation to university teaching performance.

Once the conclusions are presented, it can be noted that the present study achieved its purpose of determining the best performance model for the predictive analysis of university teaching performance, which is why it can be used as part of the procedure to improve the quality of the educational process. Because these results allow to have a relevant, reliable database that is obtained in less time compared to manual processes.

ACKNOWLEDGMENT

Thanks to the researchers who have contributed their knowledge in the development of this paper.

REFERENCES

- [1] E. Frank, M. Hall, L. Trigg, G. Holmes and I. H. Witten, "Data mining in bioinformatics using Weka," *Bioinformatics*, vol. 20, no. 15, pp. 79-81, 2004. DOI: 10.1093/bioinformatics/bth261.
- [2] T. J. Fontalvo-Herrera, E. J. Delahoz and A. A. Mendoza-Mendoza, "Data Mining Application for the Classification of High Quality Accredited Industrial Engineering University Programs in Colombia," *Technological information*, vol. 29, no. 3, pp. 89-96, 2018. DOI: <http://dx.doi.org/10.4067/S0718-07642018000300089>.
- [3] S. Bayne, "Higher education as a visual practice: seeing through the virtual learning environment," *Teaching in Higher Education*, vol. 13, no. 4, pp. 395-410, 2008. DOI:10.1080/13562510802169665.
- [4] P. Sökkhey and T. Okazaki, "Study on Dominant Factor for Academic Performance Prediction using Feature Selection Methods," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 8, pp. 492-502, 2020. DOI: 10.14569/IJACSA.2020.01110862.
- [5] E. De-La-Hoz and L. Polo, "Application of Cluster Analysis Techniques and Artificial Neural Networks in the Evaluation of the Export Potential of a Company," *Technological information*, vol. 28, no. 4, pp. 67-74, 2017. DOI: 10.4067/S0718-07642017000400009.
- [6] D. Moonsamy, N. Naicker, T. T. Adeliyi and R. E. Ogunsakin, "A Meta-analysis of Educational Data Mining for Predicting Students Performance in Programming," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, no. 2, pp. 97-104, 2021. DOI: 10.14569/IJACSA.2021.0120213.
- [7] E. F. Ruiz, E. Moreno, E. A. Carmona and L. I. Garay, "Educational Tool for Generation and Analysis of Multidimensional Modeling on Data Warehouse," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 9, pp. 261-267, 2020. DOI: 10.14569/IJACSA.2020.0110930.
- [8] D. Buenaño-Fernández, D. Gil and S. Luján-Mora, "Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study," *Sustainability*, vol. 11, no. 10, pp. 1-18, 2019. DOI: <https://doi.org/10.3390/su11102833>.
- [9] C. González, E. Elhariri, N. El-Bendary, A. Fernández and R. P. Díaz, "Machine learning based classification approach for predicting students performance in blended learning," The 1st International Conference on Advanced Intelligent System and Informatics (AIS2015), November 28-30, 2015, Beni Suef, Egypt. *Advances in Intelligent Systems and Computing*, vol. 407, pp. 47-56, 2016. DOI: https://doi.org/10.1007/978-3-319-26690-9_5.
- [10] K. H. Susheelamma and K. M. Ravikuma, "Student risk identification learning model using machine learning approach," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 5, pp. 3872-3877, 2019. DOI: <http://doi.org/10.11591/ijece.v9i5.pp3872-3879>.

- [11] R. Shadieff and M. Yang, "Review of studies on technology-enhanced language learning and teaching," *Sustainability, MDPI, Open Access Journal*, vol. 12, no. 2, p. 524, 2020. DOI: <https://doi.org/10.3390/su12020524>.
- [12] P. Ramkissoon, L. J. Belle and T. Bhurosy, "Perceptions and experiences of students on the use of interactive online learning technologies in Mauritius," *International Journal of Evaluation and Research in Education*, vol. 9, no. 4, pp. 833–839, 2020. DOI: <http://doi.org/10.11591/ijere.v9i4.20692>.
- [13] L. Medina, "Blended learning: Deficits and prospects in higher education," *Australasian Journal of Educational Technology*, vol. 34, no. 1, pp. 42–56, 2018. DOI: <https://doi.org/10.14742/ajet.3100>.
- [14] L. Soria, W. Ortega and A. Ortega, "Teaching pedagogical performance and learning of university students in the Education career," *Praxis & Know*, vol. 11, no. 27, e. 303, 2020. DOI: <https://doi.org/10.19053/22160159.v11.n27.2020.10329>.
- [15] F. E. Ceballos, J. E. Rojas, L. G. Cuba, L. P. Medina and A. R. Velazco, "Analysis of the quality of services in university centers," *University, Science and Technology*, vol. 25, no. 108, pp. 23–29, 2021. DOI: <https://doi.org/10.47460/uct.v25i108.427>.
- [16] L. Moyan and S. Yawen, "Evaluation of Online Teaching Quality of Basic Education Based on Artificial Intelligence", *International Journal of Emerging Technologies in Learning*, vol. 15, no. 16, pp. 147–161, 2020. DOI: [10.3991/ijet.v15i16.15937](https://doi.org/10.3991/ijet.v15i16.15937).
- [17] E. J. De-La-Hoz, E. De-La-Hoz and T. J. Fontalvo, "Methodology of Machine Learning for the classification and Prediction of users in Virtual Education Environments," *Technological information*, vol. 30, no. 1, pp. 247–254, 2019. DOI: <http://dx.doi.org/10.4067/S0718-07642019000100247>.
- [18] F. Makombe and M. Lall, "A Predictive Model for the Determination of Academic Performance in Private Higher Education Institutions," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 9, pp. 415–419, 2020. DOI: [10.14569/IJACSA.2020.0110949](https://doi.org/10.14569/IJACSA.2020.0110949).
- [19] H. Mushtaq, et al., "Educational Data Classification Framework for Community Pedagogical Content Management using Data Mining," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 1, pp. 329–338, 2020. DOI: [10.14569/IJACSA.2019.0100144](https://doi.org/10.14569/IJACSA.2019.0100144).
- [20] E. Ayala, R. E. Lopéz, and V. H. Menéndez, "Predictive models of academic risk in computing careers with educational data mining," *Distance Education Journal*, vol. 21, no. 66, pp. 1–36, 2020. DOI: <https://doi.org/10.6018/red.463561>.
- [21] V. Pedrero, K. Reynaldos-Grandón, J. Ureta-Achurra and E. Cortez-Pinto, "Overview of machine learning and its application in the management of emergency services," *Medical journal of Chile*, vol. 149, pp. 248–254, 2021. DOI: [10.4067/s0034-98872021000200248](https://doi.org/10.4067/s0034-98872021000200248).
- [22] S. Rajagopal, K. Siddaramappa and P. Panduranga. "Performance analysis of binary and multiclass models using azure machine learning," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 1, pp. 978–986, 2020. DOI: <http://doi.org/10.11591/ijece.v10i1.pp978-986>.
- [23] A. D. Poernomo and S. Suhajito, "Indonesian online travel agent sentiment analysis using machine learning methods," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 1, pp. 113–117, 2019. DOI: [http://doi.org/10.11591/ijeecs.v14.i1.pp113-117](https://doi.org/10.11591/ijeecs.v14.i1.pp113-117).
- [24] R. El-Shawi, S. Sakr, D. Talia, and P. Trunfio, "Big data systems meet machine learning challenges: Towards big data science as a service," *Big data research*, vol. 14, pp.1–11, 2018. DOI: [10.1016/j.bdr.2018.04.004](https://doi.org/10.1016/j.bdr.2018.04.004).
- [25] D. Pratiba and G. Shobha, "RSECM: Robust Search Engine using Context-based Mining for Educational Big Data," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 7, no. 12, pp. 39–51, 2016. DOI: [10.14569/IJACSA.2016.071206](https://doi.org/10.14569/IJACSA.2016.071206).
- [26] R. Lottering, R. Hans and M. Lall, "A Machine Learning Approach to Identifying Students at Risk of Dropout: A Case Study," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 10, pp. 417–422, 2020. DOI: [10.14569/IJACSA.2020.0111052](https://doi.org/10.14569/IJACSA.2020.0111052).
- [27] W. D. Ahmad and A. A. Bakar, "Ensemble Machine Learning Model for Higher Learning Scholarship Award Decisions," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 5, pp. 303–312, 2020. DOI: [10.14569/IJACSA.2020.01110540](https://doi.org/10.14569/IJACSA.2020.01110540).
- [28] T. C. Smith, and E. Frank, "Introducing machine learning concepts with WEKA," *In Statistical genomics, Humana Press, New York, NY*, vol. 1418, pp. 353–378, 2016. DOI: [10.1007/978-1-4939-3578-9_17](https://doi.org/10.1007/978-1-4939-3578-9_17).
- [29] Z. Ullah, F. Saleem, M. Jamjoom and B. Fakhieh, "Reliable Prediction Models Based on Enriched Data for Identifying the Mode of Childbirth by Using Machine Learning Methods: Development Study," *J Med Internet Res.*, vol. 23, no. 6, p. 28856, 2020. DOI: [10.2196/28856](https://doi.org/10.2196/28856).
- [30] E. Carpaneto, G. Chicco, R. Napoli and M. Scutariu, "Electricity customer classification using frequency-domain load pattern data," *International Journal of Electrical Power & Energy Systems*, vol. 28, no. 1, pp. 13–20, 2006. DOI: [10.1016/j.ijepes.2005.08.017](https://doi.org/10.1016/j.ijepes.2005.08.017).
- [31] U. B. Chaudhry and C. I. Phillips, "UAV Aided Data Collection for Wildlife Monitoring using Cache-enabled Mobile Ad-hoc Wireless Sensor Nodes," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, no. 10, pp. 6–17, 2021. DOI: [10.14569/IJACSA.2021.0121002](https://doi.org/10.14569/IJACSA.2021.0121002).