Supervised Learning through Classification Learner Techniques for the Predictive System of Personal and Social Attitudes of Engineering Students

Omar Chamorro-Atalaya[1], Soledad Olivares-Zegarra[2]
Facultad de Ingeniería y Gestión
Universidad Nacional Tecnológica de Lima Sur (UNTELS)
Lima- Perú

Alejandro Paredes-Soria[3], Oscar Samanamud-Loyola[4],
Marco Anton-De los Santos[5], Juan Anton-De los
Santos[6], Maritte Fierro-Bravo[7]
Facultad de Ciencias Económicas
Universidad Nacional Federico Villarreal
Lima-Perú

Victor Villanueva-Acosta[8]
Facultad de Ciencias Humanas
Universidad Autónoma del Perú
Lima-Perú

Abstract—**In this competitive scenario of the educational system, higher education institutions use intelligent learning tools and techniques to predict the factors of student academic performance. Given this, the article aims to determine the supervised learning model for the predictive system of personal and social attitudes of university students of professional engineering careers. For this, the Machine Learning Classification Learner technique is used by means of the Matlab R2021a software. The results reflect a predictive system capable of classifying the four satisfaction classes (1: dissatisfied, 2: not very satisfied, 3: satisfied and 4: very satisfied) with an accuracy of 91.96%, a precision of 79.09%, a Sensitivity of 75.66% and a Specificity of 92.09%, regarding the students' perception of their personal and social attitudes. As a result, the higher institution will be able to take measures to monitor and correct the strengths and weaknesses of each variable related to satisfaction with the quality of the educational service.**

Keywords—*Supervised learning; classification learner; predictive system; personal and social attitudes; engineering students*

I. INTRODUCTION

At present, the permanent search for educational quality is one of the main objectives of this sector [1], which is why strategies and methodologies designed to optimize student satisfaction factors [2], [3] have been implemented. Specifically, in the field of university higher education, the need to comply with quality standards in the educational service offered is evident [4].

Therefore, continuous self-evaluation of the dimensions related to student satisfaction is necessary, in order for this to improve institutional processes by identifying their strengths and weaknesses [5]-[8]. Taking into account the various influencing factors, it is complex to determine the strategic actions and decisions that correctly optimize these factors [9].

One of the factors that are related to student satisfaction is the self-perception of personal and social attitudes of university students. As indicated in [10] there is a need to evaluate the self-perception of personal and social attitudes in students, since this will generate the obligation to reflect on how the university community is contributing to social development. The importance of identifying the self-perception of the university's personal and social attitude lies in identifying what aspects need to be improved, in such a way that this leads to a significant contribution to higher education in the social context, allowing the student to meet their professional goals and personal [11], [12].

As indicated in [13], the higher educational level, has the duty to instill in university students the sense of social responsibility proper to the performance of professional activity. In this sense, the consequence of the acquisition of personal and social attitudes is a process that encompasses a significant portion in the lives of students in their passage through different organizations. Now if we develop this analysis in the context of educational virtualization, it leads us all to wonder to what extent this scenario has altered aspects related to personal and social attitudes of university students [14], [15] taking into account. As indicated in [16] that engineering students have perceived a greater change in the teaching process, due to the theoretical and practical nature of the subjects of the curriculum.

Given this, over time, information technology systems have been designed for different organizational sectors, from transactional to decision-making [17]. Within these information technology systems is the data mining tool. Data mining is conducive to the treatment of large amounts of information, and its purpose is to generate knowledge [18], [19]. Data mining uses databases, from which information is extracted in an automated way and through mathematical and statistical analysis deduces patterns and trends [20]. Data

mining makes data analysis easier, compared to traditional exploration, which, due to the large amount of data, makes this process much more complex [21].

Regarding the education sector, Educational Data Mining (EDM) is an emerging discipline that seeks to develop methods to explore data generated in the education sector, in order to achieve a better understanding of the characteristics of students and the way they learn [22]. Its development uses statistical techniques and artificial intelligence to detect patterns and anomalies in large amounts of data [23].

Within the field of artificial intelligence is Machine Learning, this learning is an automated process that extracts patterns from the data for the construction of models that allow prediction using supervised algorithms [24], [25]. One of the learning modalities that Machine Learning has is supervised learning, whose function is based on training the algorithm by granting it the questions, called characteristics, and the answers, called classes, in order that the algorithm combines them and can make predictions [26].

Within the two supervised learning techniques, is the classification, the classification algorithms look for patterns that will then allow them to classify the elements and determine which groups or classes they belong to. It should be mentioned that the values for these algorithms must be discrete values [27]. Among the classification algorithms is Kernel, which extends the regular logistic regression, used for binary classification, to deal with data that are not linearly separable [28].

Given what has been described, the need arises to design a predictive model of the personal and social attitudes of university students, which allow optimizing the services offered by the higher institution. Taking into account even more that the personal and social factors of the students are related to their academic performance [29], [30]. The research takes on even more relevance, due to the fact that, as indicated in [31], most predictive analysis research is related to primary and secondary education, so there is a small number of applications in higher education that serve the institutions as a base source for the improvement of educational quality.

In this sense, the main objective of this article is to determine the supervised learning model using the Classification Learner technique for the predictive system of personal and social attitudes of university students of professional engineering careers. To do this, it will analyze the performance metrics such as Accuracy (A), Precision (P), Sensitivity (S) and Specificity (R), to determine the most appropriate algorithm for the predictive model, also, the confusion matrix and the curve will be identified receiver operating characteristic (ROC) of the model.

The purpose of the research is to generate a significant contribution, for the taking of preventive and corrective actions that allow to comply with the quality standards of the educational service, whose improvement will be reflected in the satisfaction of the students and the academic performance.

The structure of the research development is divided into the methodology where the level of research is detailed, the participants, the data collection techniques, the validation of the collected data and the design of the supervised learning methodology through the technique classification. Next, the results and the discussion of them are presented, to finally make the conclusions of the investigation.

II. RESEARCH METHODOLOGY

A. Research Level

The research is descriptive in nature, since it focuses on determining the most optimal model of supervised learning using the Classification Learner technique for the predictive system of personal and social attitudes of university students of professional engineering careers, through the analysis of the performance metrics of the obtained algorithm (Accuracy (A), Precision (P), Sensitivity (S) and Specificity (R)).

The research starts from the identification of a problem, related to the improvement of the quality of the educational service, which is reflected in the satisfaction of the university students. Student satisfaction encompasses different dimensions, this study focuses on the self-perception of engineering students from a public university in Peru with personal and social attitudes. To do this, use of methods or tools already defined such as predictive systems through the modality of supervised learning.

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.

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 715 students. This selection criterion is part of a regulation established and approved by the public university of Peru. Due to the mandatory nature of the survey, it was possible to collect data from all participants.

TABLE I. INDICATORS THAT MEASURE THE SELF-PERCEPTION OF PERSONAL AND SOCIAL ATTITUDES

Item	Indicators
1	Take on your studies with responsibility, seriousness and dedication
2	With the pride of belonging to the university
3	With the commitment of raising the name of the university
4	With the respect you show for the authorities, teachers and administrative staff
5	With the respect that you treat your colleagues
6	With the treatment you receive from your colleagues
7	With your interest to be better every day
8	With your commitment to the surrounding society

C. Data Collection Technique and Instrument

The data collection technique is the survey, and the instrument used to collect data regarding the self-perception of engineering students about personal and social attitudes is the questionnaire, which was carried out online, at through the university's virtual platform, due to the implementation of the online teaching-learning process in the context of the health emergency. In addition, access to the online questionnaire was given through the student code of each participant.

It is necessary to indicate that the questionnaire was made up of 8 indicators (characteristics), which are detailed in Table I. Likewise, the responses to the questionnaire for the present study were transformed into a 4-level Likert scale (1: dissatisfied, 2: not very satisfied, 3: satisfied and 4: very satisfied), these levels of satisfaction, in the present analysis represent the classes of the supervised learning predictive model.

D. Reliability of the Collected Data

The validation of the collected data is carried out by means of the Cronbach's alpha coefficient through the SPSS statistical software. Table II shows the reliability result, whose value is 0.962.

E. Data Processing Design

The methodology of the research process begins with the collection of data through the online questionnaire of the self-perception of engineering students about personal and social attitudes. The responses are stored in a database, the information of which is processed by the Open Data Base Connectivity (OBDC) driver and the Matlab R2021a software. In this way, the supervised learning process begins, through the Classification Learner technique, the best Machine Learning algorithm is identified, through the evaluation of performance metrics, after which the prediction is made, which will allow the analysis of the results. In Fig. 1, the scheme of the supervisory learning methodology used is shown.

TABLE II.   CRONBACH'S ALPHA TEST

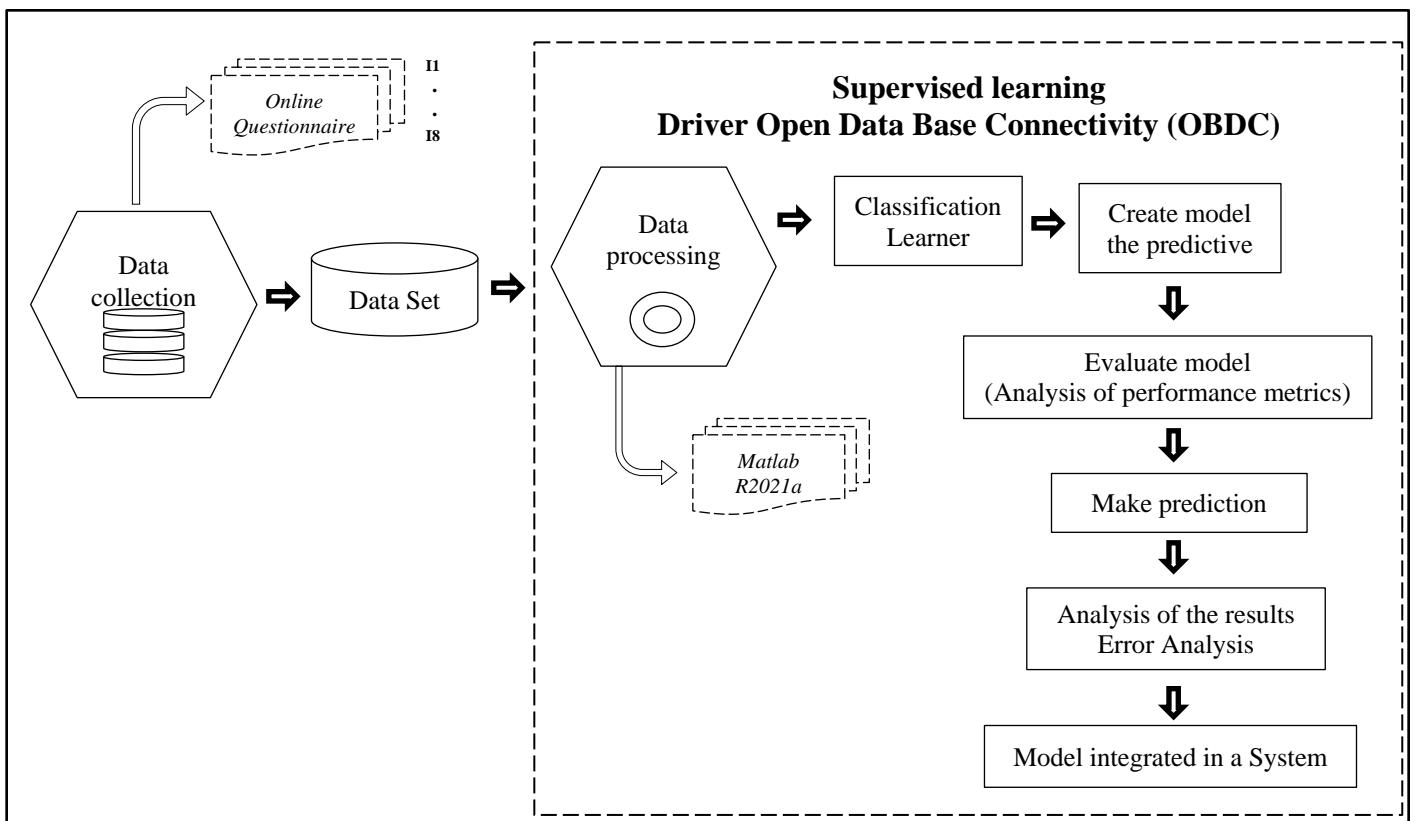| Reliability statistics	
Cronbach's alpha	*No. of elements*
0.962	8

Fig. 1.   Supervised Learning Methodology.

III. RESULT AND DISCUSSION

A. Determination of the Predictive Model

By means of the Matlab R2021a software, and through the Classification Learner technique of Machine Learning Toolbox 12.1, the best predictive model determined by the validation of the accuracy is identified, in statistical terms it is related to the bias of an estimate and is represented as the proportion of true results (true positives and true negatives) divided by the total number of cases examined (true positives, false positives, true negatives, false negatives).

In Fig. 2, the generated results are shown, which show that the Kernel algorithm: Logistic Regression Kernel, is the one that presents a better accuracy of 86.9% for the predictive system of the personal and social attitudes of engineering students This is followed by the algorithms Tree: Coarse Tree with an accuracy of 86.6%, SVM: Coarse Gaussian SVM with an accuracy of 86.4% and SVM: Linear SVM with an accuracy of 86.2%.
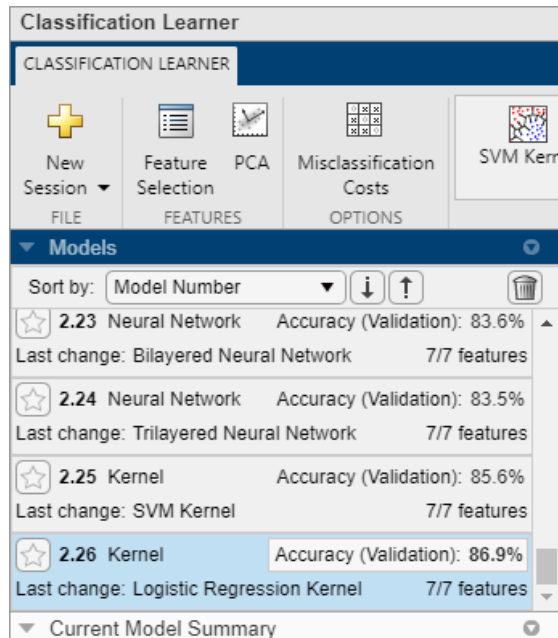

Fig. 2. Determination of the Classification Algorithm.
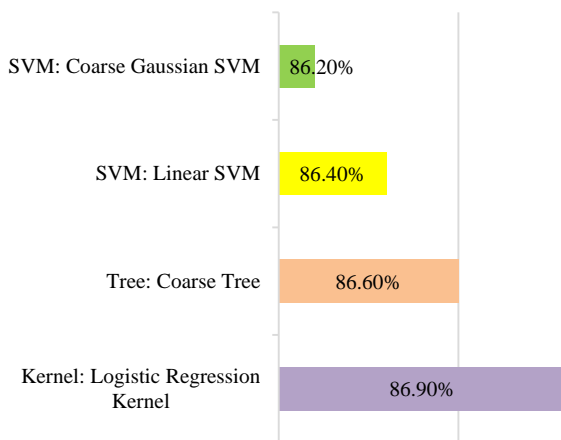

Fig. 3. Comparative Analysis of Accuracy Results.

Likewise, Fig. 3 shows the four algorithms with the highest accuracy, which identified the classification technique, where the Logistic Regression Kernel algorithm is observed as the best.

Being precision, the parameter that measures the percentage of cases that the model has hit, according to Fig. 3, it can be said that the predictive model through the Logistic Regression Kernel algorithm offers a total of 86.9% of the number of positive predictions which will be correct, that is, the value refers to how close the result of a measurement is to the true value.

B. Results of the Predictive Model Metrics

As part of the analysis of the performance metrics, we will visualize the confusion matrix of the Logistic Regression Kernel algorithm, the confusion matrix allows us to visualize the performance of a supervised learning algorithm and each column of the matrix will represent the number of predictions of each class (1: dissatisfied, 2: not very satisfied, 3: satisfied and 4: very satisfied), while each row represents the instances in the real class, in other words this analysis allows us to see what types of successes and errors our predictive system has .

In Fig. 4, the confusion matrix is shown, where the percentage of false negatives test (FNR) is displayed, also called the error rate, it is the probability that a true positive will miss it, the rate is also displayed of true positives (TPR), which measures the sensitivity metric, this metric comes to make the probability that a real positive result will be positive.

As can be seen in Fig. 4, of the 4 classes on which the predictive model acts, class 3 (satisfied) shows 89.9% sensitivity and 10.2% false negatives, this means that the predictive model has the ability to discriminate between a true positive (TP) from a false negative (FN), that is to say, there is an 89.9% capacity to be able to correctly detect satisfied students among dissatisfied students. On the other hand, class 2 (not very satisfied) shows a lower percentage of sensitivity equal to 64.4% with a percentage of false negatives of 35.6%, that is, it has the ability to detect only 35.6% of satisfied students among dissatisfied students.
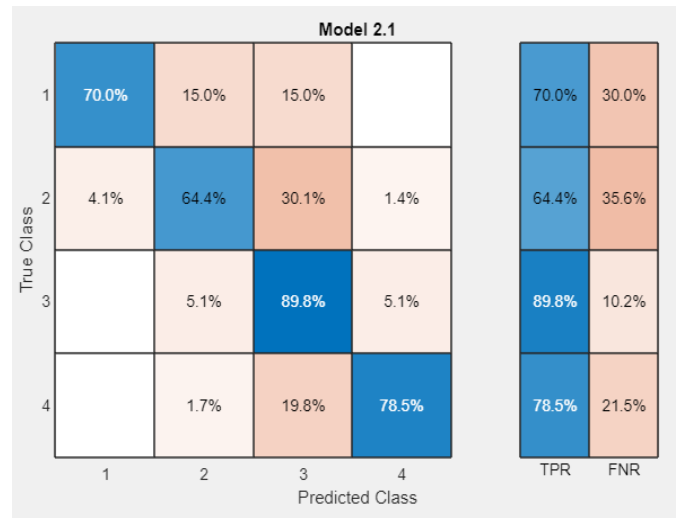

Fig. 4. Confusion Matrix based on TPR and FNR Rates.

In Fig. 5, the confusion matrix is shown, where the positive predictive value (PPV) and the false discovery rate (FDR) are displayed. It should be noted that the precision metric measures the quality of the machine learning model in classification tasks, it should be taken into account that the lower the dispersion value, the higher the precision of the model.

As can be seen in Fig. 5, the predictive model for class 3 (satisfied) shows the highest precision value, in this case it is 87.3% and a FDR percentage of 12.7%, that is, only 87.3% of students will be really satisfied with perceived personal and social attitudes, while 12.7% of examples will be wrong in the prediction.
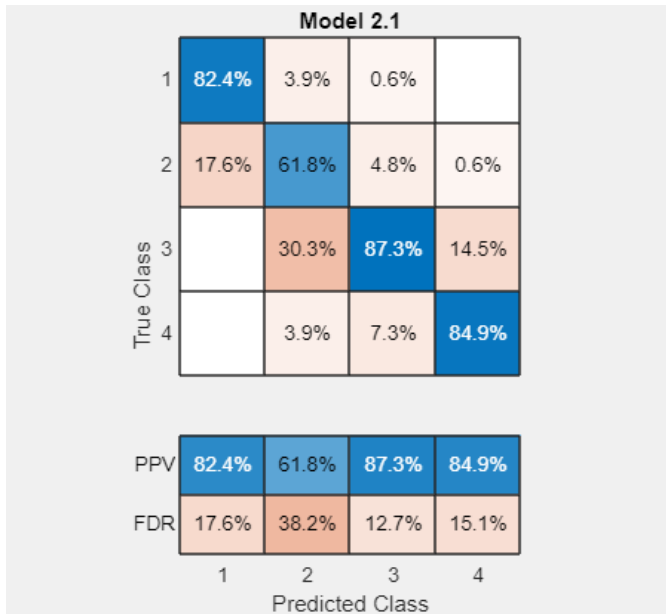

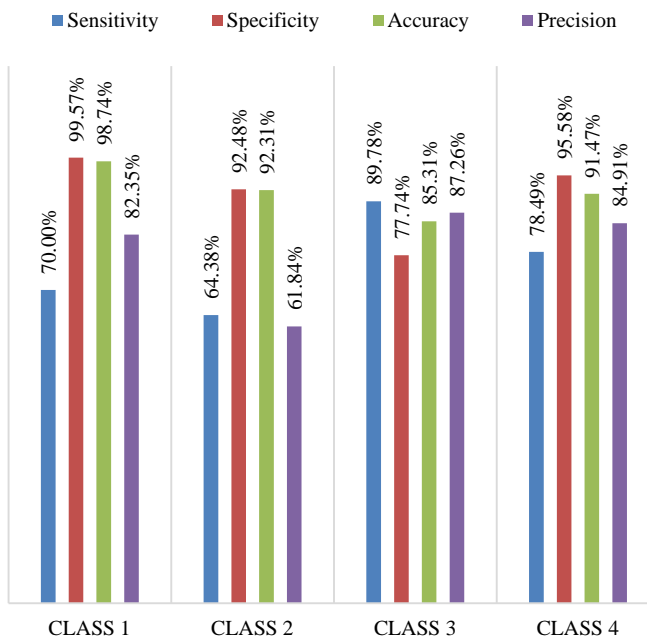Fig. 5.   Confusion Matrix based on PPV and FDR Rates.

Fig. 6.   Metrics of the Four Classes of the Predictive System.

In Fig. 6, the metrics of the predictive model of the four classes (1: dissatisfied, 2: not very satisfied, 3: satisfied and 4: very satisfied) are shown through the Logistic Regression Kernel algorithm, with this it can be said, in general, the Precision is 79.09%, the Sensitivity is 75.66%, the Specificity is 92.09% and the Accuracy is 91.96%.

Next, the analysis of the Receiver operating characteristic (ROC) curves of the Logistic Regression Kernel algorithm will be carried out, which constitutes a statistical method to determine the accuracy of the model, this test is carried out for three purposes, to determine the cut-off point of a continuous scale in the that the highest sensitivity and specificity is reached, evaluate the ability to differentiate satisfied and dissatisfied students, and compare the discriminative ability of two or more diagnoses that express their results as continuous scales. It is necessary to specify the more the value of the area on the curve (AUC), approaches 1, and the model will have a better performance and greater precision.

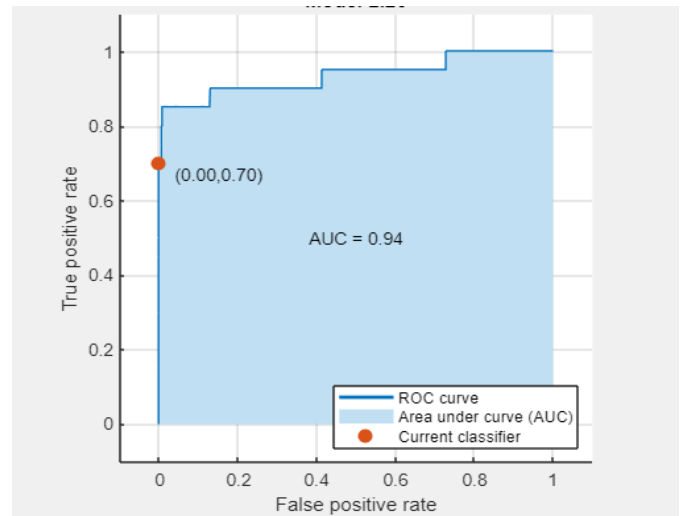In that sense, in Fig. 7, the ROC graph for class 1 (dissatisfied) is shown.

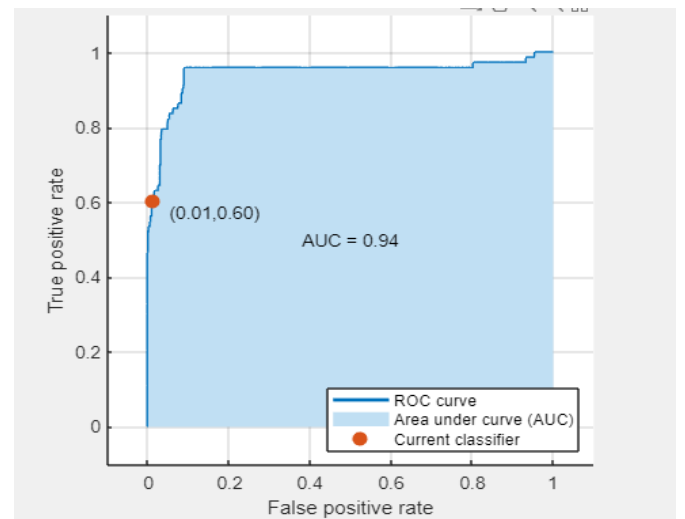Fig. 7.   ROC Charts for Class 1.
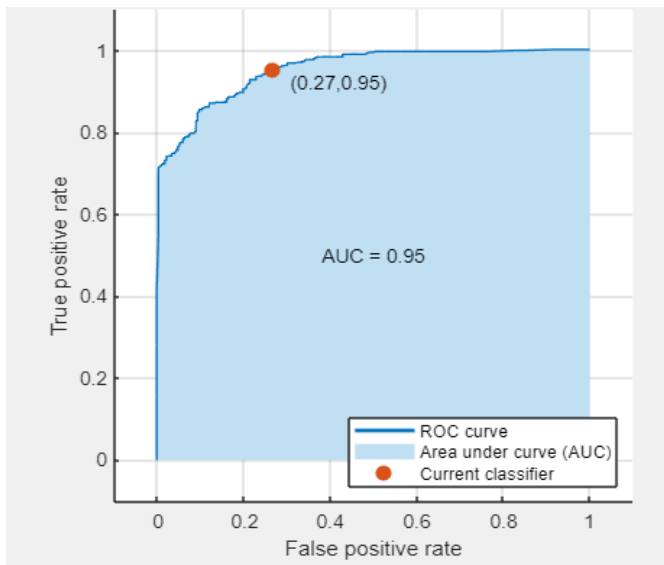

Fig. 8.   ROC Charts for Class 2.

Fig. 9.   ROC Charts for Class 3.

In Fig. 9, it is evident that for class 3, there is a sensitivity of 95%. In addition, the discrimination threshold is 0.95 for the true positive rate and 0.27 for the false positive rate, showing an optimal area value on the curve (AUC) of 0.95.

In that sense, in Fig. 10, the ROC graph for class 4 (very satisfied) is shown. For class 4, a sensitivity of 97% is evidenced. In addition, the discrimination threshold is 0.78 for the true positive rate and 0.03 for the false positive rate, showing an optimal area value on the curve (AUC) of 0.97.

Finally, it is shown in Fig. 11, the graph of parallel coordinates, used to plot multivariate data, this graph shows the relationship between the indicators of self-perception of engineering students on personal and social attitudes among the four classes of the predictive model (1: dissatisfied, 2: not very satisfied, 3: satisfied and 4: very satisfied).

In Fig. 7, it is evident that for class 1, there is a sensitivity of 94%. In addition, the discrimination threshold is 0.70 for the true positive rate and 0.00 for the false positive rate, showing an optimal area value on the curve (AUC) of 0.93.

In Fig. 8, the ROC chart for class 2 (not very satisfied) is shown.

In Fig. 7, it is evident that for class 1, there is a sensitivity of 94%. In addition, the discrimination threshold is 0.70 for the true positive rate and 0.00 for the false positive rate, showing an optimal area value on the curve (AUC) of 0.93.

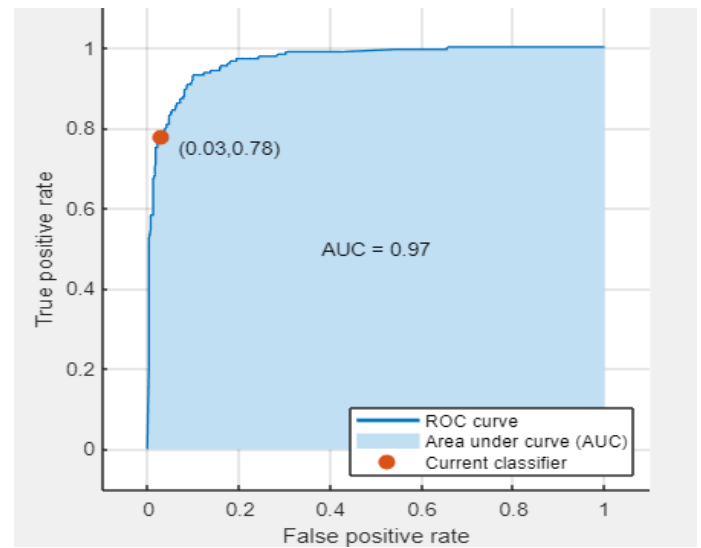In Fig. 8, the ROC chart for class 2 (not very satisfied) is shown.

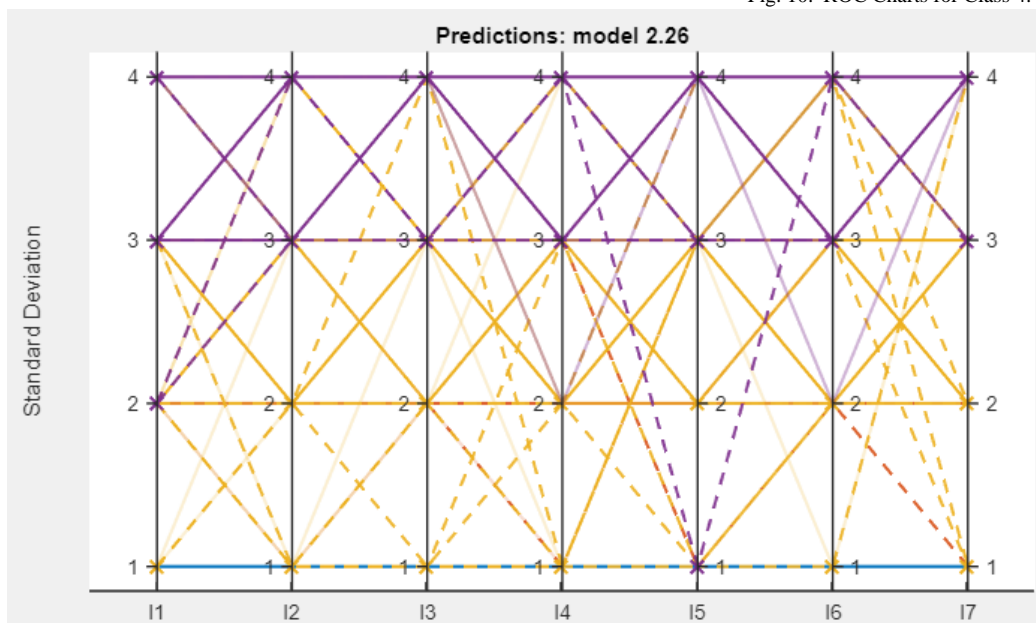Fig. 10. ROC Charts for Class 4.

Fig. 11. Parallel Coordinates Graph.

It should be noted that the data are linked to 8 descriptors related to the levels of satisfaction of the self-perception of engineering students about personal and social attitudes, during the virtual teaching-learning process, said data are ordinal qualitative types, thus defining the four classes, likewise, as part of the determination of the predictive model using the Matlab R2021a software, the data collected was used based on the first 7 indicators, which are called predictors, and the indicator of the variable under analysis (which represents the satisfaction of the self-perception of engineering students on personal and social attitudes), is defined by the final indicator (I8), because it is the one that generally encompasses the variable under analysis, for this reason in Fig. 11 only displayed from indicator I1 to indicator I7.

C. Discussion

The results of the prediction model metrics through the Logistic Regression Kernel algorithm, in relation to satisfaction with personal and social attitudes, generally present, in its 4 classes, a precision of 79.09%, a Sensitivity of 75.66%, a Specificity of 92.09% and an Accuracy of 91.96%.

Also the ROC chart for class 4 (very satisfied) shows a sensitivity of 97%, a discrimination threshold is 0.78 for the rate of true positives and 0.03 for the rate of false positives, showing an optimal area value on the curve (AUC) of 0.97.

Regarding the optimal value of the performance metrics, it is sustained that they are optimal values, based on the research of [18] where it is stated that a model is made to predict the academic performance of incoming students through machine learning, the Results showed that it is feasible to predict performance, since the model has 69% accuracy. Similarly, in [26] it is indicated that the results show a precision of 82%, therefore, it can be pointed out that the predictive model will have optimal performance when implemented.

Regarding the ROC obtained of 97% in the present investigation, in [20], the author points out that his predictive model was good since its general precision was 75.42% and an area under the ROC curve of 0.805. This is based on the established theory that the closer this value is to 1, the better the model will perform and the more accurate it is. In the same way in [23], the author points out that an AUC of 60% of 91% or 99%, represents a better performance of the classifier algorithm, these results being favorable for the investigation.

In [29] it is pointed out that in relation to the accuracy of the model for the academic performance class, the value obtained was 62.45% lower than the value obtained in the research carried out by [24] who, in their prediction model, obtained the value 73%, very close to the result obtained by [25] which was 69%.

Comparing the results with the research carried out by [31] where a lower accuracy value of 80% was obtained in the predictive system for the training data and 76% for the validity data, this study presents higher performance results. In turn, in the investigation of [27], 82.87% accuracy was obtained using the decision tree algorithm, representing a lower value than the result of our investigation.

In the study carried out by [19] a data from 914 students was used to predict their final classification (passed or failed), with a predictive model of supervised learning and the Classification Learner technique and thus obtain a model to predict the results of the students, showing that with the network algorithm (Naive de Bayes) which showed optimal precision with 71.0%, compared to four analyzed techniques (neural networks, logistic regression, decision tree and Bayesian network) with a higher percentage for the class passed and minor for failed.

IV. CONCLUSION

With the results obtained, it is evident that it is possible to apply the supervised learning methodology uses Classification Learner techniques for the predictive system of the personal and social attitudes of the students, through the graph of parallel coordinates the association and / or relationship of the indicators of the variable under analysis with the four classes of the predictive model through the Logistic Regression Kernel algorithm. The results of the prediction model metrics in relation to satisfaction with personal and social attitudes are concluded, it generally presents an optimal performance of its validation metrics in its 4 classes, with a precision of 79.09%, a Sensitivity of 75.66%, a Specificity of 92.09% and an Accuracy of 91.96%.

It is recommended to extend the line of research to others with other indicators related to student satisfaction with the educational service, because not only does it allow the higher institution to have a database or reference through a rapid and reliable classification technique for take preventive and corrective actions, to improve the quality of education, but also the predictive system influences the reduction of dropout rates and improve the academic performance of students.

REFERENCES

[1] 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.

[2] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telematics and Informatics,* vol. 37, pp. 13-49, 2019. DOI: https://doi.org/10.1016/j.tele.2019.01.007.

[3] D. Vlachopoulos, "Quality Teaching in Online Higher Education: The Perspectives of 250 Online Tutors on Technology and Pedagogy," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 16, no. 6, pp. 40-56, 2021. DOI: https://doi.org/10.3991/ijet.v16i06.20173.

[4] B. Bahati, U. Fors, P. Hansen, J. Nouri and E. Mukama, "Measuring Learner Satisfaction with Formative e-Assessment Strategies," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 14, no. 7, pp. 246-247, 2019. DOI: https://doi.org/10.3991/ijet.v14i07.9120.

[5] B. Semprún, and K. Ferrer, "Student satisfaction in a Biochemistry course: an evaluation after applying neurodidactic strategies", *San Gregorio Journal*, vol. 1, no. 38, pp. 1-14, 2020. DOI: https://dx.doi.org/rsan.v1i38.1241.

[6] I. Amoako and K. Asamoah-Gyimah, "Indicators of students' satisfaction of quality education services in some selected universities in

Ghana," *South African Journal of Higher Education*, vol. 35, no. 4, pp. 61-72, 2020. DOI: 10.20853/34-5-4252.

[7] T. Chen, L. Peng, X. Yin, J. Rong, J. Yang, and G. Cong, "Analysis of user satisfaction with online education platforms in China during the COVID-19 pandemic," *Healthc.*, vol. 8, no. 3, p. 200, 2020. DOI: 10.3390/healthcare8030200.

[8] E. Alvarado-Lagunas, J. Luyando-Cuevas and E. Piccaso-Palencia, "Perception of students towards the quality of private universities in Monterrey," *Ibero-American Journal of Higher Education*, vol. 6, no. 17, pp. 58-76, 2015. DOI: 10.1016/j.rides.2015.10.003.

[9] K. Mukhtar, K. Javed, M. Arooj, and A Sethi, "Advantages, limitations and recommendations for online learning during COVID-19 pandemic era", Pak. J. Med. Sci., vol. 36, no. COVID19-S4, pp. S27-S31, 2020, doi: 10.12669/pjms.36.COVID19-S4.2785.

[10] W. Niebles-Núñez and M. Cabarcas-Velásquez, "Social Responsibility: Training Element for University Students," *Latin American Journal of Educational Studies,* vol. 14, no. 1, pp. 257-268, 2018. DOI: https://doi. org/10.17151/rlee.2018.14.1.6.

[11] A. Bernasconi and E. Rodriguez-Ponce, "Exploratory Analysis of Perceptions on Leadership Styles, Academic Climate and the Quality of Undergraduate Education," *Form. Univ.,* vol. 11, no. 3, pp. 29-40, 2018, DOI: http://dx.doi.org/10.4067/S0718-50062018000300029.

[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] A. Ali and F. Mohammed, "Measuring quality of E-Learning and Desaire2Learn in the College of Science and Humanities at Alghat, Majma-ah University," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 9, no. 1, pp. 61-72, 2020. DOI: 10.14569/IJACSA.2018.090170.

[14] H. Hafiza and R. Ibrahim, "Distance Education during COVID-19 Pandemic: The Perceptions and Preference of University Students in Malaysia Towards Online Learning," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, no. 4, pp. 118-126, 2021. DOI: 10.14569/IJACSA.2021.0120416.

[15] C. Coman, L.G. Țîru, L. Meseșan-Schmitz, C. Stanciu, and M. C. Bularca, "Online teaching and learning in higher education during the coro-navirus pandemic: Students' perspective," *Sustainability*, vol. 12, no. 24, pp. 1-22, 2020. DOI: 10.3390/su122410367.

[16] F. Phang, et al., "Integrating Drone Technology in Service Learning for Engineering Students," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 16, no. 15, pp. 78-96, 2021. DOI: https://doi.org/10.3991/ijet.v16i15.23673.

[17] A. Ramos, M. Aldude, J. Estrada, V. Señas and L. Andrade-Arenas, "Analysis of the use of technological tools in university higher education using the soft systems methodology," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 7, pp. 412-420, 2020, DOI: 10.14569/IJACSA.2020.0110754.

[18] H. Sadiq, A. Dahan, Neama, F. M. Ba-Alwi and N. Ribata, "Educational data mining and analysis of students' academic performance using weka," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 9, no. 2, pp. 447-459, 2018. DOI: https://doi.org/10.11591/ijeecs.v9.i2.pp447-459.

[19] C. H. Menacho, "Prediction of academic performance applying data mining techniques," *Learning and Individual Scientific Annals,* vol. 78, no. 1, pp. 26-33, 2017. DOI: http://dx.doi.org/10.21704/ac.v78i1.811.

[20] R. S. Baker and K.Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3-16. DOI: https://doi.org/http://doi.ieeecomputersociety.org/10.1109/ASE.2003.1240314.

[21] E. B. Costa, B. Fonseca, M. A. Santana, F. F. De Araújo and J. Rego, "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses," *Computers in Human Behavior*, vol. 73, pp. 247–256, 2017. DOI: https://doi.org/10.1016/J.CHB.2017.01.047.

[22] B. Bakhshinategh, O. R. Zaiane, S. ElAtia and D. Ipperciel, "Educational data mining applications and tasks: A survey of the last 10 years," *Education and Information Technologies*, vol. 23, no. 1, pp. 537-553. DOI: https://doi.org/10.1007/s10639-017- 9616-z.

[23] 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.

[24] O. D. Castrillón, W. Sarache, William and S. Ruiz-Herrera, "Prediction of academic performance using artificial intelligence techniques," *University education*, vol. 13, no. 1, pp. 93-102, 2020. DOI: https://doi.org/10.4067/S0718-50062020000100093.

[25] 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, p. 2833. DOI: https://doi.org/10.3390/su11102833.

[26] R. Katarya, J. Gaba, A. Garg, Aryan and V. Verma, "A review on machine learning based student's academic performance prediction systems," *2021International Conference on Artificial Intelligence and Smart Systems (ICAIS),* pp. 254-259, Coimbatore, India: IEEE, 2021. DOI: https://doi.org/10.1109/ICAIS50930.2021.9395767.

[27] P. Sokkhey 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.2020.0110862.

[28] 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.

[29] B. Díaz-Landa, R. Meleán-Romero and W. Marín-Rodriguez, "Academic performance of students in Higher Education: predictions of influencing factors from trees decision," *Redalyc*, vol. 23, no. 3, pp. 616-635, 2021. DOI: https://doi.org/10.36390/telos233.08.

[30] M. Bravo, S. Salvo and C. Muñoz, "Profiles of Chilean students according to academic performance in mathematics: An exploratory study using classification trees and random forests," *Studies in Educational Evaluation*, vol. 44, pp. 50-59, 2015. DOI: https://doi.org/10.1016/j.stueduc.2015.01.002.

[31] M. Imran, S. Latif, D. Mehmood and M. S. Shah, "Student Academic Performance Prediction using Supervised Learning Techniques," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 14, no. 14, pp. 92–104, 2019. DOI: https://doi.org/https://doi.org/10.3991/ijet.v14i14.10310.