Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 19 No. 11 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v19i11.39887

PAPER

K-Fold Cross-Validation through Identification of the Opinion Classification Algorithm for the Satisfaction of University Students

Omar Chamorro-Atalaya[1](✉), José Arévalo-Tuesta[2], Denisse Balarezo-Mares[3], Anthony Gonzáles-Pacheco[4], Olga Mendoza-León[4], Manuel Quipuscoa-Silvestre[4], Gregorio Tomás-Quispe[5], Raul Suarez-Bazalar[6]

[1]Universidad Nacional Tecnológica de Lima Sur, Lima, Peru

[2]Universidad Nacional Federico Villarreal, Lima, Peru

[3]Universidad Privada San Juan Bautista, Lima, Peru

[4]Universidad Nacional de Trujillo, Lima, Peru

[5]Universidad César Vallejo, Lima, Peru

[6]Universidad Nacional del Callao, Lima, Peru

ochamorro@untels.edu.pe

ABSTRACT

When using machine-learning techniques to determine algorithms or ranking models that identify student satisfaction, algorithms are often trained and tested on a single data set, leading to bias in their performance metrics. This article aims to identify the best algorithm to classify the satisfaction of university students applying the K-fold cross-validation technique, comparing the error rates of the performance metrics before and after its application. The method used began with the collection of student opinions on the teaching performance of the social network Twitter during an academic semester. Then, sentiment analysis was used for data processing, through which it was possible to categorize the opinions of the students into "satisfied" or "dissatisfied." The results showed that the algorithm with the lowest error rate in its performance metric was the support vector machine (SVM). In addition, it was identified that its classification probability reached an accuracy of 91.76%. It is concluded that SVM classification using K-fold cross-validation will contribute to determining which factors associated with the teacher's didactic strategies should be improved in each class session, since traditional surveying techniques have shortcomings.

KEYWORDS

satisfaction, university students, classification algorithm, cross-validation, sentiment analysis

1 INTRODUCTION

Machine learning, like other tools linked to artificial intelligence, is proving useful in different fields, such as industry, education, medicine and the economy, since it contributes to the generation of knowledge for decision-making [1–3]. Thus, in the last two decades, the application of machine learning in the educational field has grown significantly [4–6]; this is because its main purpose is to identify classification or prediction models, based on certain particular characteristics of students [7, 8].

Chamorro-Atalaya, O., Arévalo-Tuesta, J., Balarezo-Mares, D., Gonzáles-Pacheco, A., Mendoza-León, O., Quipuscoa-Silvestre, M., Tomás-Quispe, G., Suarez-Bazalar, R. (2023). K-Fold Cross-Validation through Identification of the Opinion Classification Algorithm for the Satisfaction of University Students. *International Journal of Online and Biomedical Engineering (iJOE)*, 19(11), pp. 140–158. https://doi.org/10.3991/ijoe.v19i11.39887

Article submitted 2023-03-25. Resubmitted 2023-05-29. Final acceptance 2023-05-29. Final version published as submitted by the authors.

This allows early identification of aspects related to the educational process as a whole, contributing to improving the quality of academic service [9–11]. Among the multiple indicators that are linked to university educational quality is student satisfaction [12–14]. Being able to measure this is important, as it allows identifying which services of the educational system require action plans to improve it and obtain a competitive advantage in relation to other university institutions [15, 16]. Today, through machine-learning models, it is possible not only to identify student satisfaction, but also to predict and classify groups of students with common patterns [17–19]. However, applying machine learning entails taking into account four categories: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning [20–22]. In relation to supervised learning, it is required that the data be labeled or structured as input and output data, establishing an association between these data, and generating regression or classification patterns [23–25].

Together with machine learning, it is possible to resort to text mining through social networks to identify the satisfaction and dissatisfaction of students regarding the services they receive from the university [26–28]. One of the fields of text mining, which allows us to predict student opinions, is sentiment analysis [29, 30]. Sentiment analysis, also called opinion mining, allows classifying a series of texts by identifying the positive or negative polarity of the opinions [31–33]; as such, the algorithms or models of machine learning allow assigning a category to the opinions of the students [34]. Opinion categorization is an integral part of natural language processing (NLP), which helps to extract relevant words and retrieve information [35–37]. For the collection of opinions related to student satisfaction, the use of social networks such as Facebook, Twitter, or WhatsApp is viable [38, 39]. Thus, by using text mining or opinion mining and NLP, it is possible to identify the perception of university student satisfaction [40, 41].

There are machine-learning algorithms that use sentiment analysis such as Naive Bayes, Logistic Regression, and SVM (support vector machine) for the classification of university student satisfaction [42–44]. These are evaluated through their performance metrics [45], these being accuracy, precision, F1 score, AUC-ROC (area under the curve–receiver operating characteristic), and recall [46–48]. By comparing these metrics, it is possible to identify the machine-learning algorithm or classification model [49]. The process of training and testing the algorithm usually relies on a single data set, so it is possible to generate bias in the performance metrics of the algorithm [50]; however, through cross-validation techniques, it is possible to train, validate, and test with multiple data sets or folds [51–53]. There are several techniques to perform cross-validation, these being K-fold cross-validation, stratified K-fold, or nested cross-validation [54, 55]. Of all the aforementioned techniques, K-fold validation is the most advantageous because it uses all the data to train and validate, obtaining more representative results a priori and with less bias or risk of error [56]. Thus, by applying the K-fold technique to determine student satisfaction, the identification of the algorithm's performance metrics is guaranteed with levels of precision and accuracy without bias [57].

This article has as its research question: What is the algorithm for classifying university student satisfaction that presents the best performance metric, after applying the K-fold cross-validation technique? The method that will be used will take as its starting point the collection of student opinions on teacher performance from the social network Twitter. These data will be collected during one academic semester, and subsequently sentiment analysis will be used for data processing through the

NLTK (natural language toolkit) and Vader (valence-aware dictionary and sentiment reasoner) libraries of Python, with which the opinions of the students will be categorized as satisfied and dissatisfied. Likewise, for the application of the K-fold cross validation, the Python software libraries will be used.

2 LITERATURE REVIEW

Machine learning has the purpose of getting computers to process large volumes of data, seeking to generate classification or prediction algorithms [58]. Based on the type of learning, these algorithms are classified as unsupervised, supervised, or reinforcement learning [59]. In unsupervised learning, the main task is to identify groups of unlabeled data with a common feature, called clusters [60], while, in supervised learning, the dataset under analysis is labeled as input data and output data, thus generating models or prediction algorithms [61]. Among the most common algorithms, we have the regression algorithm, the decision-tree algorithm, the SVM, and Naive Bayes [62]. These algorithms are chosen according to their performance indicators or metrics, these being precision, accuracy, sensitivity (recall), F1 score, and AUC [63]. Accuracy is the metric that helps to know how exact or close the result is to the true value, providing information about the possible errors that can be found in the classification [64]. Equation (1) shows the expression to determine the accuracy of the classification algorithm, where accuracy depends on the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) [65].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Likewise, sensitivity or recall represents the proportion of correctly classified positive cases [66]. Equation (2) shows the expression to determine sensitivity [67].

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

AUC represents the probability of correctly classifying a randomly chosen positive class rather than a randomly chosen negative one [66]. Equation (3) shows the expression to determine the area under the curve.

$$AUC = 1 - specificity = \frac{FP}{TN + FP} \tag{3}$$

Another of the tools used in this study is NLP, whose purpose is to create computer systems that understand, process, and generate a natural language similar to that used by human beings [68]. It comprises a set of techniques and strategies that guarantee that systems interpret and process human language (spoken, symbolic, and written) [69]. Its application implies the use of the following steps: implementation of libraries that allow the extraction of text, tokenization, conversion of words to lowercase text, filtering of special characters, elimination of punctuation marks, verification of letter contents, elimination of spaces between words, correction of words, and change of word for its corresponding synonym according to the context [70]. Table 1 shows the list of studies in which machine-learning techniques, sentiment analysis, NLP, classification algorithms, and K-fold cross-validation have been used. For each research study reviewed, the title of the manuscript is specified, as well as the results obtained. These results will be useful for the discussion section of the results found in the K-fold cross-validation study by identifying the opinion classification algorithm for classifying the satisfaction of university students.

Table 1. Previous studies

References	Manuscript Title	Identified Results
[71]	Enhancing sentiment analysis of textual feedback in the student faculty evaluation using machine learning techniques	The researchers identified the performance metrics of the algorithms for classifying student satisfaction with respect to teacher performance, for which they used sentiment analysis techniques with the purpose of categorizing the opinions of the students, working with a single training sample and testing the algorithm. It was identified that the Random Forest algorithm showed high performance. However, when compared with the performance of the mixed n-gram algorithm composed with SVM, the latter was better.
[72]	Perceiving university students' opinions from Google app reviews	The researchers identified the classification algorithm on the opinions of university students regarding the use of Google applications, in which they compared performance metrics of machine-learning models such as Random Forest, SVM, KNN (K- Nearest Neighbors), Naive Bayes and Logistic Regression. They identified that the algorithm with the best performance metrics was the SVM.
[73]	Design of a predictive model on the dropout of an electrical and electronic engineering student at the University of the Andes using machine learning techniques	The researchers developed a related investigation on the identification of the algorithm for predicting the classification of university student dropout, in which they applied the K-fold technique in order to guarantee that the results obtained by each of the machine-learning models were truly independent of the selection of data for training and testing, with a value of K equal to 4.
[74]	Analysis of academic performance using machine learning techniques with assembly methods	This study used the K-fold cross-validation technique with a value of K equal to 10, with the purpose of compensating for bias effects in the training data set, to obtain a classification model on academic performance of university students.
[75]	Predictive classification model based on automated learning for the early detection of potential drop-out college students	The research seeks to identify the predictive classification model based on machine learning applied to the university environment. To validate the performance of the classification algorithms, K-fold of K folds was used, where K took a value equal to 10. The result was that the algorithm with the highest performance was Random Forest.
[76]	Business intelligence model to analyze social media information	This research seeks to develop a tool that identifies customer comments regarding opinions generated in social networks regarding the performance of an organization, for which it makes use of supervised learning techniques. As a result, the algorithm with the best precision turned out to be the SVM.
[77]	An extensive study of sentiment analysis techniques: A survey	In this study, the textual data-mining technique is used, in which it seeks to classify the opinions of customers. This study determined the accuracy of hybrid lexicon-focused machine learning performance.

Cross-validation is an important sample-processing method for modern statistics and is widely used in machine-learning models [78, 79]. There are several cross-validation techniques, one of which is the K-fold cross-validation [80]. Figure 1 shows the cross-validation process with K equal to 10, which consists of the random division of a data set into 10 subgroups, or folds [81]. One of these folds is the test group, and the other remaining K-1 subgroups are called the training group. Training and validation are performed K times, using a different subgroup as the validation set in each iteration and the rest of K-1 as the training set [82].

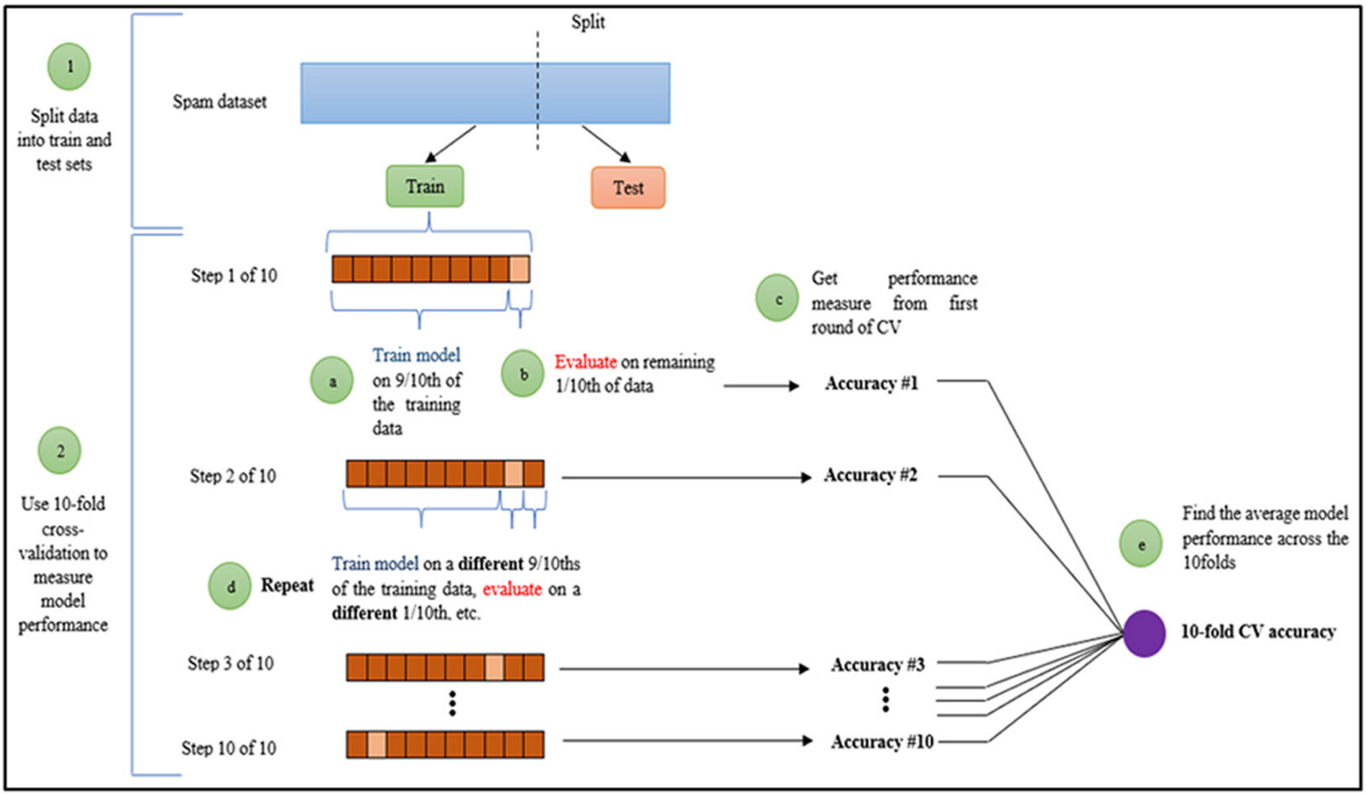

Fig. 1. Steps to perform the K-fold technique

In each of the K iterations of this type of validation, an error calculation is performed. In this regard, in [83] it is indicated that the final result is obtained from the arithmetic mean of the K values of errors obtained, according to equation (4).

$$E = \frac{1}{k}\sum_{k=1}^{k} E_i \tag{4}$$

3 METHODOLOGY

3.1 Data-collection stage

The method used has as its starting point the data-collection stage, which was made up of student opinions that were generated on the Twitter social network. A tweet was created weekly, which contained an open question addressed to the students in which they were asked: What do you think about the teacher's performance in the class session? Previously, an explanation was made to the students about the purpose of their opinions and what factors are linked with respect to the satisfaction of the teaching performance. In this way, fifteen tweets were generated throughout the academic semester, resulting in collecting 254 responses or opinions from students. These opinions were grouped weekly and stored in a database with a CSV (comma-separated values) extension that could be read and processed by Python software.

3.2 Opinion-processing stage through NLP

The second stage consisted of the processing of opinions through the NLP, which was developed through the use of the NLTK and Vader libraries from Python. This stage, in turn, is made up of three sub-stages. The first sub-stage, called "data pre-processing," consisted of cleaning the texts written by the students, eliminating duplicate words and extra spaces between words, and converting all texts to lower case. The second sub-stage consisted of the "weighting of sentiment," that is, quantitative values between –1 and 1 were assigned to each opinion. The third sub-stage consisted of the "categorization of the feelings" contained in the opinions of the students as "dissatisfied" and "satisfied"; this was achieved using a vectorization technique such as TF-IDF (term frequency– inverse document frequency). That is, those opinions with a weight between -1 and 0, called opinions with a negative polarity, were categorized as dissatisfied, while those opinions with a weight between 0 and 1, called opinions with a positive polarity, were categorized as satisfied. It should be noted that the opinions whose weighting turned out to be zero were not considered in the study because they do not contribute to the increase in the percentage of satisfied or dissatisfied students.

3.3 Identification stage of the classification algorithm

The third stage consisted of the identification of the classification algorithm, and like the previous stages, it was supported by Python machine-learning libraries. The first sub-stage consisted of the "designation of data for training and testing," establishing 70% of the data collected for training and 30% for testing; This is because, generally, for the use of the K-fold cross-validation technique, data are used in the proportion of 70% and 30% or 80% and 20% [62]. The evaluation algorithms were the SVM, Logistic Regression, Naive Bayes, and Decision Tree; using of all of them, it was possible to identify their performance metrics (accuracy, F1 score, and AUC-ROC).

The second sub-stage consisted of the "identification of the performance metrics," but now applying the stratified type K-fold cross-validation technique with a value of K equal to 10; this is because the amount of data collected (254 opinions) was small. This validation was repeated five times for each algorithm evaluated. Finally, the third sub-stage consisted of the "comparison of the performance metrics and identification of the classification algorithm" of student satisfaction regarding teacher performance. Figure 2 shows the method used to identify the algorithm for classifying student satisfaction through K-fold cross-validation.

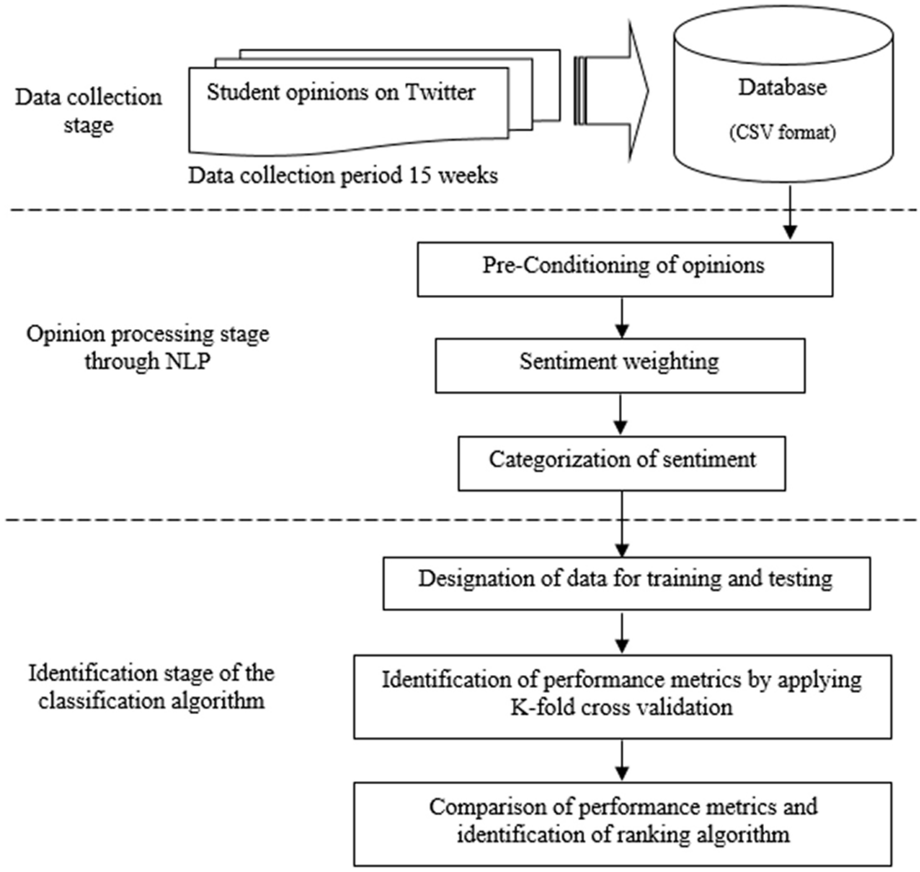

Fig. 2. Method used to identify the classification algorithm by K-fold cross-validation

4 RESULTS AND DISCUSSION

This section considers the stages of the method defined for the identification of the university student satisfaction classification algorithm through K-fold cross-validation. The first stage was data collection. Table 2 shows the number of tweets obtained per class week, in which it was possible to collect 288 tweets or opinions from the students. It should be noted that of the 288 opinions collected, 34 opinions obtained neutral polarity, so the tweets to be processed for the evaluation was reduced to 254.

Table 2. Tweets collected by class week

Opinion Collection Week	Tweets with Positive Polarity	Tweets with Negative Polarity	Tweets with Neutral Polarity	Total Tweets Per Week
Week 1	26	0	1	27
Week 2	17	1	4	22
Week 3	18	0	3	21
Week 4	16	1	1	18
Week 5	8	2	4	14
Week 6	15	0	3	18

(Continued)

Table 2. Tweets collected by class week *(Continued)*

Opinion Collection Week	Tweets with Positive Polarity	Tweets with Negative Polarity	Tweets with Neutral Polarity	Total Tweets Per Week
Week 7	16	5	0	21
Week 8	11	7	6	24
Week 9	10	2	1	13
Week 10	15	2	3	20
Week 11	19	3	2	24
Week 12	15	1	1	17
Week 13	15	1	1	17
Week 14	13	1	1	15
Week 15	14	0	3	17

In relation to the stage corresponding to the processing of opinions through sentiment analysis, Table 3 shows the categorization of the 254 opinions generated by students regarding teaching performance. It was found that 90.55% of opinions were categorized as satisfied, while 9.45% were categorized as dissatisfied.

Table 3. Categorization of opinions

Number	Tweet	Sentiment
0	Sorry teacher, I couldn´t get into the class but…	Satisfied
1	I was not able to attend the first class. But…	Satisfied
2	The session presented was interesting and moti…	Satisfied
3	Interesting session. Understandable to famil…	Satisfied
4	The class session seemed very precise and I re…	Satisfied
…	…	…
249	The final presentation of the article was very…	Satisfied
250	It was possible to put together a good introdu…	Satisfied
251	Today was the investigation of our academic wo…	Satisfied
252	The elaboration of the academic work throughout…	Satisfied
253	Today was the presentation of our TA. I partic…	Satisfied

In order to predict the classification as "satisfied" or "dissatisfied" of the subsequent opinions on student satisfaction generated in other academic semesters, different machine-learning models were evaluated to obtain their performance metrics and identify which of them predicted the classification with greater precision. In this research, the Support Vector Machine, Logistic Regression, Naive Bayes and Decision Tree models were trained and tested. Figure 3 shows the results of the precision and the F1 score as performance metrics of the classification algorithm. At first glance, it could be assumed that the algorithm that shows the best performance metric is the Decision Tree algorithm; however, these results were obtained with a single training dataset and a single test dataset, distributed in a proportion of 70% and 30%, respectively. Another aspect to highlight is the low values of the F1 score metric, which shows that there is an imbalance in the data collected.

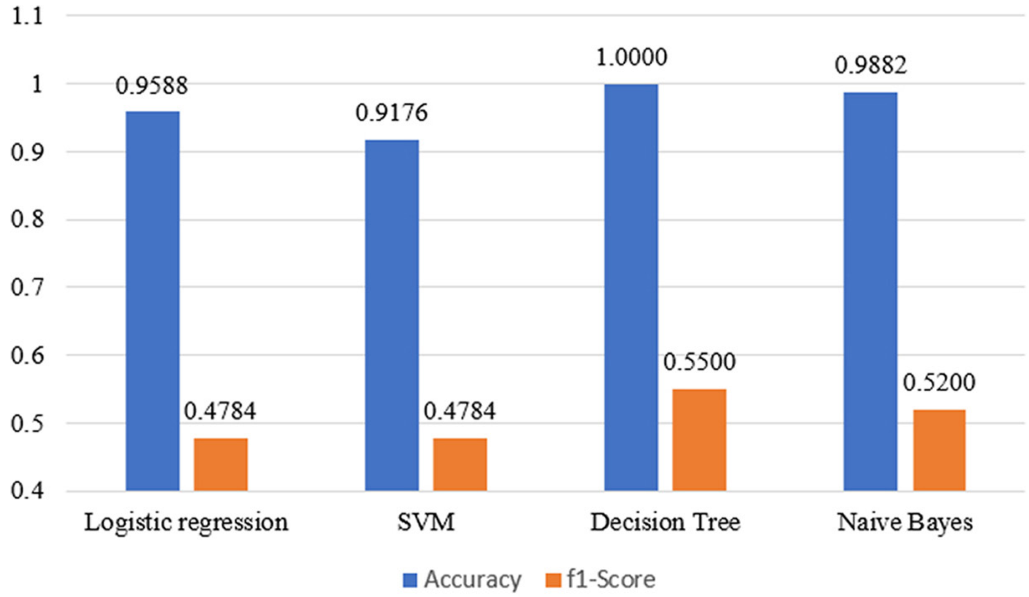

Fig. 3. Performance metrics identified before the cross-validation technique

When identifying high values of the accuracy metric and low values of the F1-score metric, it is necessary to validate the results obtained through techniques such as K-fold cross-validation. For purposes of applying cross-validation in this research, a value of K equal to 10 was used, with five cases of validation for each iteration of K; this was to achieve greater precision in the results of performance metrics. Bravo et al. [64] point out that by using a K-fold cross-validation with a value of K equal to 10, it is possible to compensate for the effects of bias in the training data set. Table 4 shows the results of the K-fold cross-validation applied to the SVM algorithm, which shows the results of the accuracy of the algorithm. It can be seen that this value changes for different samples or training data (fold), as well as for each validation case. To describe the result obtained after the validation technique, the average value of all the folds in all the cases analyzed was obtained, resulting in an average value of accuracy equal to 0.9176.

Table 4. Results of cross-validation of the SVM algorithm

Classification Algorithm: SVM	First Case	Second Case	Third Case	Fourth Case	Fifth Case
1-fold	0.9412	0.8235	0.9412	0.9412	0.8824
2-fold	0.8824	0.8824	0.9412	0.8824	0.8824
3-fold	1.0000	0.9412	0.8824	1.0000	0.9412
4-fold	1.0000	0.8824	1.0000	0.8824	0.8824
5-fold	0.9412	0.9411	0.9412	0.9412	1.0000
6-fold	0.7647	0.8824	0.8824	0.9412	0.9412
7-fold	0.9412	1.0000	1.0000	0.9412	0.9412
8-fold	0.8824	0.9412	0.8824	0.8824	0.8235
9-fold	0.9412	1.0000	0.9412	0.8235	0.9412
10-fold	0.8824	0.8824	0.8824	0.9412	0.9412

Table 5 shows the results of the cross-validation for the Logistic Regression algorithm. The average value of the accuracy reaches a result identical to that of the SVM algorithm.

Table 5. Results of the cross-validation of the logistic regression algorithm

Classification Algorithm: Logistic Regression	First Case	Second Case	Third Case	Fourth Case	Fifth Case
1-fold	0.9412	0.8235	0.9412	0.9412	0.8824
2-fold	0.8824	0.8824	0.9412	0.8824	0.8824
3-fold	1.0000	0.9412	0.8824	1.0000	0.9412
4-fold	1.0000	0.8824	1.0000	0.8824	0.8824
5-fold	0.9412	0.9411	0.9412	0.9412	1.0000
6-fold	0.7647	0.8824	0.8824	0.9412	0.9412
7-fold	0.9412	1.0000	1.0000	0.9412	0.9412
8-fold	0.8824	0.9412	0.8824	0.8824	0.8235
9-fold	0.9412	1.0000	0.9412	0.8235	0.9412
10-fold	0.8824	0.8824	0.8824	0.9412	0.9412

Table 6 shows the results of the K-fold cross-validation, for the Decision Tree algorithm. The accuracy metric of the algorithm oscillates from a minimum value of 07059 to a maximum value of 1; the average value of said metric is equal to 0.8929.

Table 6. Results of the cross-validation of the decision tree algorithm

Classification Algorithm: Decision Tree	First Case	Second Case	Third Case	Fourth Case	Fifth Case
1-fold	0.9412	0.8235	0.9412	0.9412	0.8824
2-fold	0.8235	0.8235	0.8824	0.8235	0.8824
3-fold	0.9412	0.8235	0.8235	1.0000	0.8824
4-fold	0.9412	0.8824	1.0000	0.8824	0.7647
5-fold	0.9412	0.9412	0.8824	0.9412	0.7059
6-fold	0.8235	0.9412	0.9412	0.8235	0.9412
7-fold	0.9412	1.0000	0.8235	0.8824	0.9412
8-fold	0.8824	0.8824	0.8824	0.8824	0.8235
9-fold	0.9412	1.0000	0.8235	0.8824	0.8824
10-fold	0.8824	0.9412	0.9412	0.9412	0.8824

The fourth algorithm to which the K-fold cross-validation was applied was Naive Bayes. T average value of the accuracy metric is 0.8906. Table 7 shows the results of the validation for each fold and for the five repetitions.

Table 7. Results of the cross-validation of the Naive Bayes algorithm

Classification Algorithm: Naive Bayes	First Case	Second Case	Third Case	Fourth Case	Fifth Case
1-fold	0.9412	0.8235	0.9412	0.9412	0.8235
2-fold	0.7647	0.8824	0.8235	0.8235	0.8824
3-fold	0.9412	0.9412	0.9412	1.0000	0.9412
4-fold	1.0000	0.8235	1.0000	0.8824	0.8235
5-fold	0.9412	0.9412	0.9412	0.9412	0.9412
6-fold	0.7647	0.7647	0.7059	0.8824	0.9412
7-fold	0.8824	1.0000	0.8235	0.9412	0.9412
8-fold	0.9412	0.8824	0.8824	0.8824	0.8235
9-fold	0.8824	0.9412	0.8824	0.7647	0.8824
10-fold	0.8824	0.8824	0.8824	0.8824	0.9412

Figure 4 shows the results obtained after applying the K-fold cross-validation. In shows that the algorithms with the best accuracy in predicting the classification of student satisfaction are the SVM algorithm and Logistic Regression, both with accuracy values equal to 0.9176, higher than those achieved by the other two algorithms.

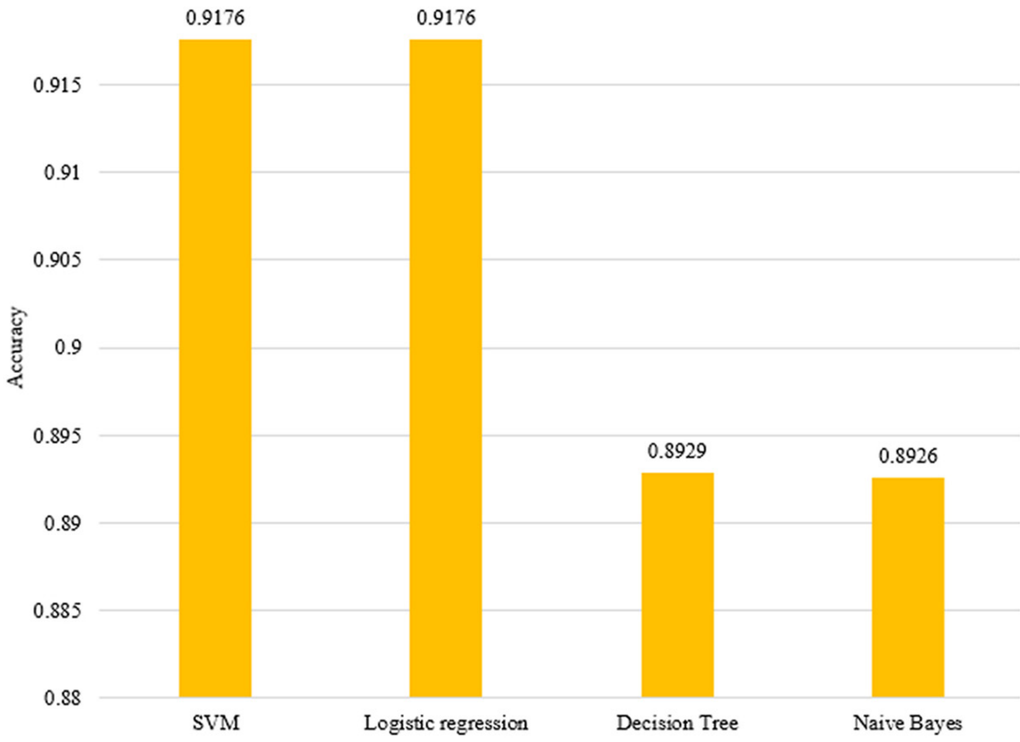

Fig. 4. Results of the accuracy metric after applying K-fold cross-validation

In order to carry out a more exhaustive analysis regarding the performance metrics of the classification algorithms and to be able to identify the one that best performs the classification, the AUC-ROC metric was determined. Its value represents the sensitivity of the algorithm; that is, when the AUC-ROC is approximately 0.5, the model does not have the capacity to discriminate between a satisfied and dissatisfied class, and when the AUC-ROC value is close to 0, the algorithm classifies the satisfied class as an unsatisfied class and vice versa. Therefore, to establish that the algorithm has optimal sensitivity, it must be close to 1. Sucapuca [65] points out that in order to compare the models in a general way, an ROC curve is prepared, which represents the result of the sensitivity indicators; however, the ROC curve is a graphic representation, so in order to be compared, it is necessary to calculate the area that each represents, which is called AUC. Figure 5 shows the representation of the ROC curves for the different classification algorithms analyzed, obtained from the use of machine-learning libraries of the Python software. Figure 5(a) corresponds to the ROC curve of the Logistic Regression algorithm, in which the sensitivity value represented by the AUC value turned out to be, on average for the 10 folds of the cross-validation, equal to 0.83. In Figure 5(b) the corresponding value using SVM algorithm was 0.84, in Figure 5(c), the corresponding value for the Decision Tree algorithm was 0.69, and in Figure 5(d), the corresponding value for the Naïve Bayes algorithm was 0.60.

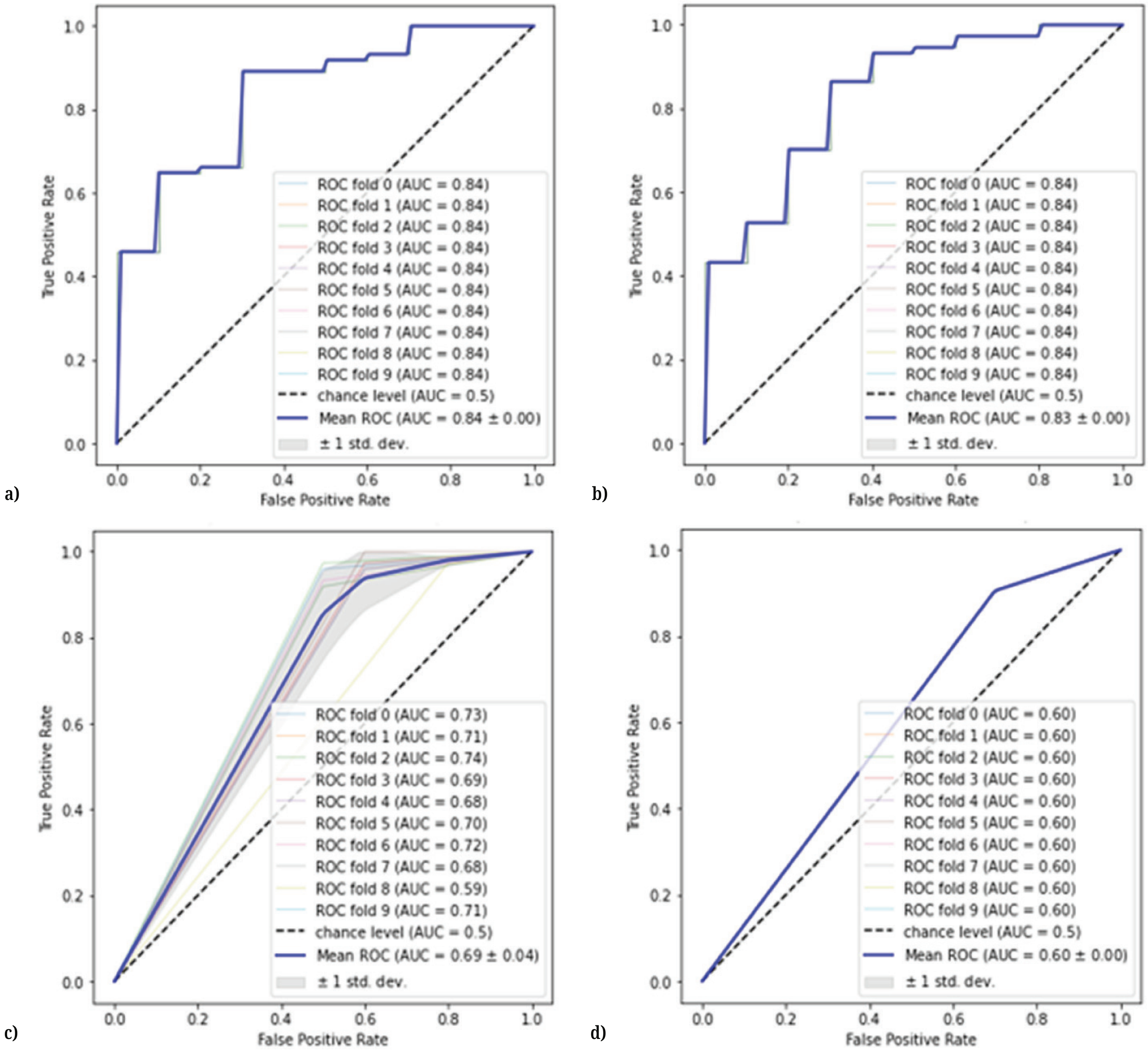

Fig. 5. ROC curves for the **(a)** Logistic Regression, **(b)** SVM, **(c)** Decision Tree, and **(d)** Naïve Bayes

Figure 6 shows the comparative results of the accuracy metric before and after the application of cross-validation. In the case of the Logistic Regression algorithm, before the application it had an accuracy of 0.9588, and after the application it reached a value of 0.9176. In the case of the SVM algorithm, before the application it had a value of 0.9176, and after the application it maintained the same value. In the case of the Decision Tree algorithm, before the application it had a value of 1, and after the application it reached a value of 0.8929. In the case of the Naive Bayes algorithm, before the application it had an accuracy of 0.9882, and after the application of the cross-validation it reached a value of 0.8926. Of all these results, the only algorithm that showed no bias in the performance metric "accuracy" was the SVM algorithm.

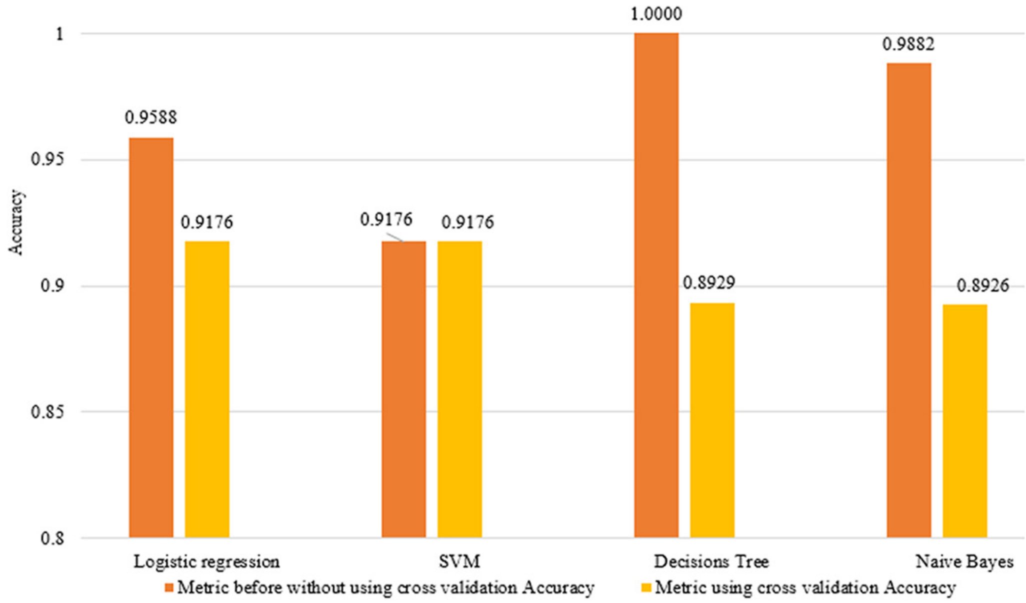

Fig. 6. Comparison of the accuracy before and after the application of the K-fold cross-validation

According to the results of both the accuracy metric and the sensitivity metric, of the four algorithms analyzed, the classification algorithms with the best performance were the SVM and Logistic Regression. In general, both algorithms had similar performance. In the case of the SVM algorithm, it predicted the classification with an accuracy of 91.76%, while the proportion of true positives that were correctly predicted reached 84%. Pacol and Palaoag [61] were able to determine that the SVM algorithm reached an accuracy value of 0.98. However, they did not use the cross-validation technique, and, as evidenced in this research, high values of the performance metrics are not a sign that the algorithm will really make a correct prediction of the classification. This is due to over-fitting when using a single data set for training or due to class imbalance in the collected data. This leads to the presence of bias in the results, which makes it necessary to use cross-assessment. Rajan and Mishra [62] concluded that when analyzing the classification algorithm using the K-Fold cross-validation technique, they were able to determine that the classification algorithm with the best accuracy turned out to be the SVM algorithm, reaching a value of 93.41%. This supports what has been demonstrated in this research because it is necessary to apply cross-validation techniques to achieve greater precision with respect to the performance metrics of the algorithm. Giving further support to the results obtained in this research, Kurnia [72] used K-fold cross-validation to identify the classification algorithm with the best performance and determined that the SVM algorithm reached an accuracy of 78.99%, evidencing that the classification matrices performance varied in relation to the training and testing data sets.

5 CONCLUSION

From the results obtained, it was determined that, by applying the technique of sentiment analysis and automatic learning with the purpose of identifying the algorithm that classifies the opinions of the students with the best performance, the performance metrics of all the algorithms evaluated experienced significant changes when submitting them to a K-fold cross-validation procedure. This is because the classification algorithm, when the cross-validation technique is not applied, is trained and tested with only a single dataset, generating bias in its performance metrics. In addition, when comparing which algorithm shows a higher performance after applying cross-validation, it was determined that the SVM algorithm reached a level of precision of 91.76% and sensitivity (AUC-ROC) of 84%. This means that the SVM algorithm predicts the ranking of student opinions between the satisfied and dissatisfied classes with a high level of accuracy. In other words, this algorithm is sensitive to discriminate both types of opinions. It is concluded that SVM classification using K-fold cross-validation will contribute to determining which factors associated with the teacher's didactic strategies should be improved in each class session, since traditional surveying techniques have shortcomings.

6 STUDY LIMITATIONS

The study presents as a limitation the lack of university regulations that make the continuous participation of students sustainable during all class sessions. Currently the regulations of the National Technological University of Lima Sur (UNTELS) in Peru indicate that a satisfaction survey can be administered only once in the academic semester in which the teaching performance, among other aspects, is evaluated. It also indicates that its application will be made after fifteen weeks of the start of classes. This current scenario means that students do not have a culture of permanent participation in the evaluation processes of the satisfaction of the academic service received. For this reason, when we carried out this investigation, only 254 opinions were obtained and only from students of the automatic process control course of the mechanical and electrical engineering professional career program. Future research could be carried out with a larger population made up of students from other professional schools.

7 REFERENCES

[1] Y. Kumar, K. Kaur, and G. Singh, "Machine learning aspects and its applications to-wards different research areas", in *2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, 2020, pp. 150–156. https://doi.org/10.1109/ICCAKM46823.2020.9051502

[2] D. Santos, L. Dallos, and P. A. Gaona-García, "Motion tracking algorithms using artificial intelligence and machine learning techniques," *Technological Information*, vol. 31, no. 3, pp. 23–38, 2020. https://doi.org/10.4067/S0718-07642020000300023

[3] J. R. Nogales and M. Benalcázar, "Real-Time hand gesture recognition using the leap motion controller and machine learning," in *Conference Proceedings UTMACH*, 2019, pp. 823–835. https://doi.org/10.1109/LA-CCI47412.2019.9037037
[4] I. Naqa and M. J. Murphy, "What is machine learning?" in *Machine Learning in Radiation Oncology*. Springer Cham, 2015, pp. 3–11. https://doi.org/10.1007/978-3-319-18305-3_1
[5] E. Cruz, M. González, and J. C. Rangel, "Machine learning techniques applied to performance evaluation and dropout prediction of university students, a review," *Technlogical Prism*, vol. 13, no. 1, pp. 77–87, 2022. https://doi.org/10.33412/pri.v13.1.3039
[6] I. Majeed and S. Naaz, "Current state of art of academic data mining and future vision," *Indian J. Comput. Sci. Eng.*, vol. 9, no. 2, pp. 49–56, 2018. https://doi.org/10.21817/indjcse/2018/v9i2/180902026
[7] J. F. V. García, "Forecast model of academic performance of students in the courses of the basic studies program of the Ricardo Palma University using Machine Learning algorithms," MS. thesis, Ricardo Palma University, Lima, Peru, 2019.
[8] M. M. Afshin, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*, 2nd ed, Cambridge, Massachusetts London: The MIT Press, 2020.
[9] D. Zhang, "Analysis of university management model of national higher education institutions based on machine learning algorithm," *Mob. Inf. Syst.*, vol. 2022, pp. 1–7, 2022. https://doi.org/10.1155/2022/4553185
[10] Z. M. Rodriguez, "Machine learning process to determine the social demand for IT professional jobs," *Industrial Data Journal*, vol. 25, no. 2, pp. 275–300, 2022. https://doi.org/10.15381/idata.v25i2.21643
[11] D. E. M. Cervera, "Forecast model in higher education in specific tests of engineering, bachelor's degree and mathematical scientific thinking in saber pro applying machine learning," MS. thesis, University Francisco Jose de Calda, Bogotá, Colombia, 2021.
[12] M. I. L. Tarazona, R. S. Perales-Flores, and M. L. Marcelo-Armas, "Student satisfaction of the teaching management of a public University in Huánuco-Perú," *Scientific Gazette*, vol. 6, no. 2, pp. 80–91, 2020. https://revistas.unheval.edu.pe/index.php/gacien/article/view/785
[13] R. F. Díaz-Camacho, J. L. R. Muñoz, I. A. E. Díaz, and U. I. R. Miranda, "Student satisfaction in virtual education: an international systematic review," *Chakiñan J. Soc. Sci. Humanit.*, no. 16, pp. 177–193, 2022.
[14] R. P. Leyva, "Academic satisfaction of nursing students at a Public University in Mexico," *Electronic Journal on Technology, Education and Society*, vol. 4, no. 7, pp. 1–14, 2017.
[15] E. R. P. Mazariego, E. G. S. Pérez, and J. F. G. Rodríguez, "Validation of the model to measure university student satisfaction with the academic services received," *Oper. Res.*, vol. 41, no. 3, pp. 472–481, 2020.
[16] E. O. Carhuaz, "Academic satisfaction of university students in the framework of virtual education," *Scientific Journal of Social Communication*, no. 2, pp. 16–24, 2020.
[17] E. G. Surdez-Pérez, M. D. C. Sandoval-Caraveo, and C. L. Lamoyi-Bocanegra, "Student satisfaction in the assessment of university educational quality," *Education and Educators*, vol. 21, no. 1, pp. 9–26, 2018. https://doi.org/10.5294/edu.2018.21.1.1
[18] I. M. K. Ho, K. Y. Cheong, and A. Weldon, "Predicting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques," *PLoS One*, vol. 16, no. 4, pp. 1–27, 2021. https://doi.org/10.1371/journal.pone.0249423
[19] A. R. Almasri, N. A. Yahaya, and S. S. Abu-Naser, "Instructor performance modeling for predicting student satisfaction using machine learning - preliminary results," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 19, pp. 5481–5496, 2022.
[20] M. Mohammed, M. B. Khan, and E. B. M. Bashier, *Machine Learning: Algorithms and Applications*. Florida: CRC Press, 2016, Ch. 2. https://doi.org/10.1201/9781315371658

[21] A. Gron, *Hands-On Machine Learning with Scikit Learn Keras and Tensorflow*, 2nd ed, Sebatopol, CA: O'Reilly Media, 2019.

[22] C. A. B. Cruz, "Comparative analysis between supervised learning and semi-supervised learning for the classification of volcanological seismic signals of the cotopaxi volcano," B.S. thesis, National Polytechnic School, Quito, Ecuador, 2020.

[23] F. A. González, "Machine learning models in rheumatology," *Rev. Colomb de Reumatol.*, vol. 22, no. 2, pp. 77–78, 2015. https://doi.org/10.1016/j.rcreu.2015.06.001

[24] 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," *Rev. Med. Chile.*, no. 149, pp. 248–254, 2021. https://doi.org/10.4067/s0034-98872021000200248

[25] P. E. G. Fuentes, "Application of machine learning algorithms to predict profit per customer based on google analytics metrics," BS. thesis, University of Valladolid, Valladolid, Spain, 2019.

[26] A. B. Urbina-Nájera, J. C. Camino-Hampshire, and R. Cruz Barbosa, "University dropout: prevention patterns through the application of educational data mining." *RELIEVE - Electronic Journal of Educational Research and Evaluation*, vol. 26, no. 1, pp. 1–19, 2020. https://doi.org/10.7203/relieve.26.1.16061

[27] E. A. Franco, R. E. L. Martínez, and V. H. M. Domínguez, "Predictive models of academic risk in computing careers with educational data mining," *J Distance Educ.*, vol. 21, no. 66, pp. 1–36, 2021.

[28] A. A. R. Gómez, and A. V. Ibáñez, "Data mining application in the virtual education," *Colombian Journal of Advanced Technologies*, vol. 1, no. 29, pp. 92–98, 2017.

[29] J. F. G. Évora, "Applications of text mining in the 2019 national transparency survey: a new analysis alternative for perception surveys," MS. thesis, Costa Rica University, San José, Costa Rica, 2020.

[30] A. F. G. Viera, "Machine learning techniques used for text mining," *Library Research*, vol. 31, no. 71, pp. 103–126, 2017.

[31] J. S. Mosquera, "Comparison between machine learning algorithms in the implementation of sentiment analysis for texts in Spanish," BS. Thesis, University Francisco José de Caldas, Bogotá, Colombia, 2021.

[32] P. González-Flores, C. R. Maya, and I. R. Z. Sierra, "Here to stay after the pandemic: sentiment analysis of higher education teachers at UNAM," in *2021 XVI Latin American Conference on Learning Technologies (LACLO)*, 2021, pp. 144–150. https://doi.org/10.1109/LACLO54177.2021.00021

[33] C. Benavides-Morales and B. Herrera-Flores, "Educational data mining: sentiment analysis in a university domain during the pandemic," *J. Lat. Am. Commun. Res.,* no. 151, pp. 217–236, 2023. https://doi.org/10.16921/chasqui.v1i151.4760

[34] J. R. Saura, A. Reyes-Menéndez, and P. Palos-Sánchez, "A feeling analysis in twitter with machine learning: capturing sentiment from #BlackFriday Offers," *Spaces Magazine*, vol. 39, no. 42, pp. 1–11, 2018.

[35] S. Ranjan and S. Mishra, "Comparative sentiment analysis of app reviews," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2020, pp. 1–7. https://doi.org/10.1109/ICCCNT49239.2020.9225348

[36] F. M. Ramos and J. I. Velez, "Integration of natural language processing techniques through web services," B. S. thesis, National University of the Center of the Province of Buenos Aires, Buenos Aires, Argentina, 2016.

[37] M. Hernández and J. Gómez, "Natural language processing applications," *Polytechnic Journal*, vol. 32, no. 1, pp. 87–96, 2020.

[38] A. Vizcaino-Verdu and I. Aguaded, "Sentiment analysis on Instagram: polarity and subjectivity of children's accounts," *ZER - J. Commun. Stud.*, vol. 25, no. 48, pp. 213–229, 2020. https://doi.org/10.1387/zer.21454

[39] H. M. Vilca-Apaza, Y. M. Mamani-Mamani, and F. S. Gutiérrez, "Social Networks and its relationship with the level of academic performance in university educational students in the Andean Region of Peru," *Journal of Human Sciences, Social Theory and Critical Thinking*, no. 15, pp. 137–154, 2022.

[40] O. Chamorro-Atalaya et al., "sentiment analysis through twitter as a mechanism for assessing university satisfaction," *Indones. J. Electr. Eng. Comput. Sci.,* vol. 28, no. 1, pp. 430–440, 2022. https://doi.org/10.11591/ijeecs.v28.i1.pp430-440

[41] F. A. Lovera and Y. Cardinale, "Sentiment analysis in Twitter: A comparative study," *Scientific Journal of Systems and Informatics*, vol. 3, no. 1, pp. 1–17, 2023.

[42] P. A. Pauli, "Sentiment analysis: comparison of predictive algorithms and methods using a Spanish lexicon," BS. thesis, Technological Institute of Buenos Aires – ITBA, Buenos Aires, Argentina, 2019.

[43] H. M. Z. Valdivies, "Computational model to albert early the level of achievement of the graduation profile of university engineering students," PhD. dissertation, University of Trieste, Trieste, Italy, 2021.

[44] V. D. G. Vera, "Methodology to predict student performance in virtual distance university courses," PhD. dissertation, National university of Colombia, Medellín, Colombia, 2021.

[45] R. Borja-Robalino, A. Monleón-Getino, and J. Rodellar, "Standardization of performance metrics for machine and deep learning classifiers," *Iberian Journal of Information Systems and Technologies*, no. E30, pp. 184–196, 2020.

[46] J. J. C. Pacheco, "Classification of medical images with deep learning techniques," BS. thesis, University of Extremadura, Badajoz, Spain, 2012.

[47] M. L. R. Lara, J. P. A. Chuquimarca, and A. R. C. Guanoluisa, "Comparative analysis of machine learning algorithms for the detection of pests in representative crops of the Ecuadorian highlands," BS. thesis, Central University of Ecuador, Quito, Ecuador, 2021.

[48] G. P. González, "Detection of fraudulent credit card transactions through the use of machine learning models," BS. thesis, University of the Andes, Bogotá, Colombia, 2021.

[49] A. M. Langan and W. E. Harris, "National student survey metrics: where is the room for improvement?" *High Education*, vol. 78, pp. 1075–1089, 2019. https://doi.org/10.1007/s10734-019-00389-1

[50] S. Arlot and A. Celisse, "A survey of Cross-Validation procedures for model selection," *Stat Surv*, vol. 4, pp. 40–79, 2010. https://doi.org/10.1214/09-SS054

[51] L. A. N. Romero, "Automatic categorization of responses applying supervised classification algorithms to the analysis of student responses to a series of test-type questions," MS. thesis, University of the Basque Country, Basque Country, Spain, 2016.

[52] N. A. R. Murillo, "Cross validation analysis under different noise conditions," MS. thesis, National Technological of Mexico in Celaya, Celaya, Mexico, 2019.

[53] D. C. C. Martínez and A. G. P. Indacochea, "Determination of authorship of texts in Spanish through stylometric analysis of frequently used words and cross validation for machine learning," BS. thesis, Guayaquil University, Guayaquil, Ecuador, 2022.

[54] R. L. Almagro, "Design of a system based on Support Vector Machines for the detection of brain tumors through hyperspectral images," BS. thesis, Higher Technical School of Engineering and Telecommunication Systems, Madrid, Spain, 2017.

[55] J. A. G. Mejía, I. D. Jiménez, and J. J. S. Acosta, "Parfase model approach to support decision making on diseases diagnosis," *Science and Engineering*, vol. 38, no. 2, pp. 113–130, 2017.

[56] L. Pérez-Planells, J. Delegido, J. P. Rivera-Caicedo, and J. Verrelst, "Analysis of Cross-Validation Methods for robust obtaining of biophysical parameters," *Remote Sensing Journal*, vol. 2015, no. 44, pp. 55–65, 2015. https://doi.org/10.4995/raet.2015.4153

[57] S. Y. A. Salinas and L. F. P. Gonzabay, "Lexical complexity prediction system implementing machine learning and neural networks to reduce barriers to reading comprehension in university students," BS. thesis, Guayaquil University, Guayaquil, Ecuador, 2021.

[58] V. D. Gil-Vera and C. Seguro-Gallego, "Machine Learning applied to software development performance analysis," *Polytechnic Journal*, vol. 18, no. 35, pp. 128–139, 2022. https://doi.org/10.33571/rpolitec.v18n35a9

[59] S. Rodríguez-Tapia and J. Camacho-Cañamón, "Supervised Machine Learning methods in textual classification according to the degree of specialization," *Journal of Philological Studies*, no. 35, pp. 1–28, 2018.

[60] I. C. P. Verona and L. A. García, "A review on unsupervised learning of distance metrics," *Cuban Journal of Informatics Sciences*, vol. 10, no. 4, pp. 43–67, 2016.

[61] A. Z. Hernández, G. A. G. Rosales, H. J. J. Santiago, and M. M. Lee, "Performance metrics to evaluate machine learning in the classification of petroleum images using convolutional neural networks," *Science Latin Multidisciplinary Scientific Journal*, vol. 6, no. 5, pp. 4624–4637, 2022.

[62] C. A. R. Vásconez, "Application of Machine Learning algorithms to predict student dropout in first and second semester students at Public Universities in Ecuador," MS. thesis, Technical University of Ambato, Ambato, Ecuador, 2023.

[63] J. M. G. Carchipulla and W. R. S. Morocho, "Development of a virtual assistant using natural language processing and machine learning for communication with applicants for postgraduate studies of the Salesian Polytechnical University using messaging channels," BS. thesis, Salesian Polytechnical University Cuenca Headquarters, Cuenca, Ecuador, 2022.

[64] J. O. Y. Pedraza, "Study of the state of the art on the prediction of university dropout using machine learning," BS. thesis, Salesian Polytechnic University Quito Headquarters, Quito, Ecuador, 2023.

[65] J. A. G. Lozada and E. Y. S. Arana, "Data exploration & book's system recommendation," *Illuminate Student Research Journal*, vol. 12, no. 1, pp. 35–46, 2020.

[66] A. M. Grisales-Aguirre and C. J. Figueroa-Vallejo, "Modeling of topics applied to the analysis of the paper of automatic learning in systemic revisions," *Research, Development and Innovation Journal*, vol. 12, no. 2, pp. 279–292, 2022. https://doi.org/10.19053/20278306.v12.n2.2022.15271

[67] J. C. B. Andrade, "Applications of natural language processing," *Between Science and Engineering*, vol. 16, no. 31, pp. 7–8, 2022.

[68] M. I. O. Velázquez, "Natural language processing for the Analysis of Self-Reflexive Emotional Writing (EEA) in women victims of violence," MS. thesis, National Technological of Mexico, Cuernavaca, Morelos, México, 2022.

[69] A. D. J. Alfaro and J. V. D. Ospina, "Systematic literature review: Machine Learning techniques," *Active Notebook*, vol. 13, no. 1, pp. 113–121, 2021.

[70] J. A. H. Castro, "Application of Machine Learning in financial credit risk management: A systematic review," *Interfaces*, no. 15, pp. 160–178, 2022.

[71] C. Pacol and T. D. Palaoag, "Enhancing sentiment analysis of textual feedback in the student faculty evaluation using machine learning techniques," *Eur. J. Eng. Sci. Tech.*, vol. 4, no. 1, pp. 27–34, 2021. https://doi.org/10.33422/ejest.v4i1.604

[72] S. Ranjan and S. Mishra, "Perceiving university students' opinions from google app reviews," *Concurr Comput.*, vol. 34, no. 10, pp. 1–25, 2022. https://doi.org/10.1002/cpe.6800

[73] J. R. Molina, A. F. Zambrano, and L. F. Giraldo, "Design of a predictive model on the dropout of an electrical and electronic engineering student at the University of the Andes using machine learning techniques," *University of the Andes*, pp. 1–12, 2022.

[74] L. E. C. Bravo, H. J. F. López, and E. R. Trujilllo, "Analysis of academic performance using machine learning techniques with assembly methods," *Redepi Newsletter Journal*, vol. 10, no. 13, pp. 171–190, 2021. https://doi.org/10.36260/rbr.v10i13.1737

[75] J. T. Sucapuca, "Predictive classification model based on automated learning for the early detection of potential drop-out college students," BS. thesis, National University of San Agustin de Arequipa, Arequipa, Perú, 2021.
[76] S. P. F. Kurnia, "Business intelligence model to analyze social media information," *Procedia. Comput. Sci.*, vol. 135, pp. 5–14, 2018. https://doi.org/10.1016/j.procs.2018.08.144
[77] L. R. Ali, B. N. Shaker, and S. A. Jebur, "An extensive study of sentiment analysis techniques: A survey," *Al-Kadhum 2nd International Conference on Modern Applications of Information and Communication Technology*, 8–9 December 2021, Baghdad, Iraq.
[78] I. C. Pinar, "Application of classification techniques to cancer detection," BS. thesis, Sevilla University, Sevilla, Spain, 2019.
[79] L. T. Chabes, "Comparative Analysis of machine learning techniques to detect credit card fraud," B.S. thesis, Catholic University of Saint Paul, Arequipa, Peru, 2022.
[80] S. D. Romero and W. S. Mosquera, "Selection of candidates for surveys using machine learning techniques," BS. thesis, Guayaquil University, Guayaquil, Ecuador, 2021.
[81] M. J. A. M. Fritz, "Regularization methods for the selection of variables applied to the prediction of the risk of suffering motor dysfunction in active older adults of the city of Valdivia," MS. thesis, University of Concepcion, Concepción, Chile, 2018.
[82] J. J. B. Picazo "Microcontroller based presence detection system for panning cameras using artificial neural networks," BS. thesis, Malaga University, Málaga, Spain, 2017.
[83] K. Ashtray, "What is exact way to do k-fold validation?" *Cross Validated*, Apr. 02, 2018. [Online]. Available: https://stats.stackexchange.com/q/338044 [Accessed Mar. 22, 2023].

8 AUTHORS

Omar Chamorro-Atalaya is a research professor at the Faculty of Engineering and Management of the National Technological University of Lima Sur (UNTELS), Lima, Peru (email: ochamorro@untels.edu.pe).

José Arévalo-Tuesta is a research professor at the Faculty of Economic Sciences of the Federico Villarreal National University (UNFV), Lima, Peru (email: jarevalotu@unfv.edu.pe).

Denisse Balarezo-Mares is a professor at the Law School of the San Juan Bautista Private University (UPSJB), Lima, Peru (email: denisse.balarezo@upsjb.edu.pe).

Anthony González-Pacheco is a professor at the Faculty of Education and Communication Sciences of the National University of Trujillo (UNT), Lima, Peru (email: agonzalesp@unitru.edu.pe).

Olga Mendoza-León is a professor at the Faculty of Education and Communication Sciences of the National University of Trujillo (UNT), Lima, Peru (email: omendozal@unitru.edu.pe).

Manuel Quipuscoa-Silvestre is a professor at the Faculty of Education and Communication Sciences of the National University of Trujillo (UNT), Lima, Peru.

Gregorio Tomás-Quispe is a professor at the Faculty of Health Sciences of the César Vallejo University (UCV), Lima, Peru (email: gtomas@ucvvirtual.edu.pe).

Raul Suarez-Bazalar is a professor at the Faculty of Administrative Sciences of the National University of Callao (UNAC), Lima, Peru (email: rsuarezb@unac.edu.pe).