Ride Revista Iberoamericana para la Investigación y el Desarrollo Educativo
ISSN 2007 - 7467

https://doi.org/10.23913/ride.v16i32.2835

Scientific articles

Ejercicios escolares para la predicción del rendimiento académico de estudiantes universitarios con técnicas de aprendizaje automático combinadas

School exercises for predicting university students' academic performance using combined machine learning techniques

Exercícios escolares para prever o desempenho acadêmico de estudantes universitários usando técnicas combinadas de aprendizado de máquina

Andrés Rico Páez
Instituto Politécnico Nacional, México
aricop.ipn@gmail.com
https://orcid.org/0000-0002-6450-318X

Nora Diana Gaytán Ramírez
Instituto Politécnico Nacional, México
nora_diana@hotmail.com
https://orcid.org/0000-0002-5159-9194

Resumen

El propósito de este estudio es proponer una metodología para desarrollar modelos predictivos del rendimiento académico de estudiantes mediante ejercicios académicos realizados en clase y utilizando técnicas de aprendizaje automático combinadas conocidas como técnicas de voto mayoritario y de apilamiento. Se recabaron datos de 250 estudiantes universitarios de México acerca de sus evaluaciones de ejercicios escolares para elaborar los modelos y se obtuvieron métricas de desempeño con validación cruzada. Posteriormente, se aplicaron los modelos construidos a 108 estudiantes de un ciclo posterior del mismo curso y se calcularon sus métricas.

Los resultados obtenidos con validación cruzada muestran que la técnica de apilamiento que tiene en la segunda fase la técnica k vecinos más cercanos tiene una mayor exactitud (69.2%).

Cuando se predice el rendimiento académico de 108 estudiantes a partir de los modelos desarrollados, la exactitud más alta se obtiene con la técnica de apilamiento que tiene en la segunda fase la técnica k vecinos más cercanos con un valor de 74.1%. La información obtenida se recopiló en un 17% de avance temporal en el curso facilitando la detección temprana de estudiantes con problemas escolares para que los profesores realicen intervenciones oportunas y mejoraren su desempeño. Es habitual que los profesores recaben las evaluaciones de los ejercicios académicos sin requerir utilizar otras herramientas más complejas de recopilación de información lo que favorece utilizar este tipo de metodologías para construir modelos predictivos.

Palabras clave: técnicas de aprendizaje automático combinadas, apilamiento, voto mayoritario, modelos predictivos, educación superior.

Abstract

The purpose of this study is to propose a methodology for developing predictive models of student academic performance using academic exercises completed in class and combined machine learning techniques known as majority voting and stacking. Data were collected from 250 university students in Mexico regarding their assessments of school exercises to develop the models, and performance metrics were obtained through cross-validation. Subsequently, the constructed models were applied to 108 students in a later semester of the same course, and their metrics were calculated.

The results obtained through cross-validation show that the stacking technique with the k-nearest neighbors' method in the second phase has the highest accuracy (69.2%). When predicting the academic performance of 108 students using the developed models, the highest accuracy is obtained with the stacking technique that includes the k-nearest neighbors' method in the second phase, with a value of 74.1%. The information obtained was collected 17% of the way through the course, facilitating the early detection of students with academic difficulties so that teachers can intervene promptly and improve their performance. It is common for teachers to collect assessments of academic exercises without needing to use more complex data collection tools, which favors the use of this type of methodology for building predictive models.

Keywords: combined machine learning techniques, stacking, majority voting, predictive models, higher education.

Resumo

O objetivo deste estudo é propor uma metodologia para o desenvolvimento de modelos preditivos do desempenho acadêmico de estudantes, utilizando exercícios acadêmicos realizados em sala de aula e técnicas de aprendizado de máquina combinadas, conhecidas como votação majoritária e empilhamento (stacking). Os dados foram coletados de 250 estudantes universitários no México, referentes às suas avaliações de exercícios escolares, para o desenvolvimento dos modelos. As métricas de desempenho foram obtidas por meio de validação cruzada. Posteriormente, os modelos construídos foram aplicados a 108 estudantes em um semestre posterior da mesma disciplina, e suas métricas foram calculadas. Os resultados obtidos por meio da validação cruzada mostram que a técnica de empilhamento com o método k-vizinhos mais próximos (k-NN) na segunda fase apresentou a maior acurácia (69,2%). Ao predizer o desempenho acadêmico dos 108 estudantes com base nos modelos desenvolvidos, a maior acurácia foi obtida com a técnica de empilhamento que inclui o método k-NN na segunda fase, com um valor de 74,1%. As informações obtidas foram coletadas em 17% do curso, facilitando a detecção precoce de estudantes com dificuldades acadêmicas, permitindo que os professores intervenham prontamente e melhorem seu desempenho. Os professores geralmente coletam avaliações de exercícios acadêmicos sem a necessidade de ferramentas de coleta de dados mais complexas, o que favorece o uso desse tipo de metodologia para a construção de modelos preditivos.

Palavras-chave: técnicas combinadas de aprendizado de máquina, empilhamento, votação majoritária, modelos preditivos, ensino superior.

Date Received: September 2025 **Date Accepted:** February 2026

Introduction

In recent years, there have been significant technological advancements, such as mobile devices, the internet, and mobile phone networks, impacting various areas. One area that has been greatly influenced by technological development is education. The growth in information and communication technologies has facilitated learning activities, both in knowledge acquisition and in the collection of user records. In educational institutions, digital tools have improved methods for accessing and storing student information, such as grades, attendance, and credits earned, among many other attributes (Salas *et al.*, 2019). Information processing and analysis allow us to understand data and develop potential solutions to problems in education that may influence teaching and learning processes. These analyses allow us to identify similarities, differences, and patterns in the data, which can contribute to the development of educational policies and improve pedagogical practices (Zambrano *et al.*, 2024). Among the analyses of educational records used to improve teaching processes is the prediction of students' academic performance using machine learning techniques (Del Carpio, 2024).

The prediction of academic performance in university students has increased due to the interest of researchers and educational institutions in improving pedagogical strategies and identifying early on students at risk of failing who require some type of academic support or intervention (Zambrano *et al.*, 2024). Machine learning techniques have been used to predict academic performance (Vargas & Prieto, 2024). These techniques allow the construction of predictive models from a set of records called training data, with the aim of predicting and validating them with another set of data known as test data. There are studies in this line of research, such as the one conducted by Gil & Quintero (2023), in which the authors used data from students who participated in four online university courses. The authors built predictive models with different machine learning techniques based on seven predictor academic variables. Villarrasa *et al*. (2024) performed academic performance prediction models using a machine learning technique known as decision tree employing 14 predictor variables.

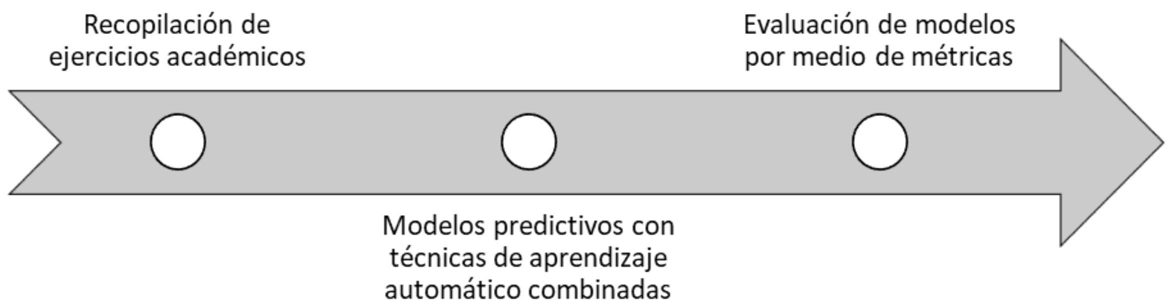
Currently, there are studies that employ predictive models of academic performance developed with combined machine learning techniques (Contreras *et al* ., 2021), that is, they combine techniques as if they were a single approach to improve predictions. In Mexico, there has been little progress in developing predictive models with these types of techniques due, among other reasons, to a lack of knowledge of these methodologies. This has led to a certain lag compared to other countries in the processing of academic data, with the potential benefit of improving teaching. Therefore, this research poses the following research questions: How can predictive models of the academic performance of students in Mexico be developed using in-class exercises and combined machine learning techniques? How can the developed predictive models be evaluated? Thus, the objective of this article is to develop a methodology for constructing predictive models of the academic performance of Mexican students at the end of a course using in-class exercises with combined machine learning techniques and to propose metrics for evaluating these models.

Ride Revista Iberoamericana para la Investigación y el Desarrollo Educativo
ISSN 2007 - 7467

Methodology

The methodology used in this research consists of collecting information from 250 students at a public university in Mexico regarding their grades obtained in five academic exercises for a mathematics course during the first few weeks of the semester. This information is used to create predictive models of course pass or fail, that is, of their academic performance at the end of the semester. This type of tool helps professors identify students most likely to fail the course and, consequently, implement interventions to help these students catch up in the early stages of the semester. Figure 1 schematically presents the methodology used .

Figure 1. Schematic of the methodology used in the research.



Source: Own elaboration

The exercises were conducted during the first three weeks of the 18-week course. Participants were informed that the collected data would be used solely for educational research purposes. Furthermore, to protect confidentiality and for ethical reasons , their names were not used; instead, they were assigned an identifier, as machine learning techniques do not require them.

In this study, exercise evaluations were scored on a scale of 0 to 10. They were classified as passing (6 or higher, ≥ 6), failing (less than 6, < 6), and exercises not submitted by students for evaluation (unsubmitted). Each student's final course grade was either pass or fail. Part of the structure of the data collected from the students is presented in Table 1.

Table 1. Structure of data collected in the research.

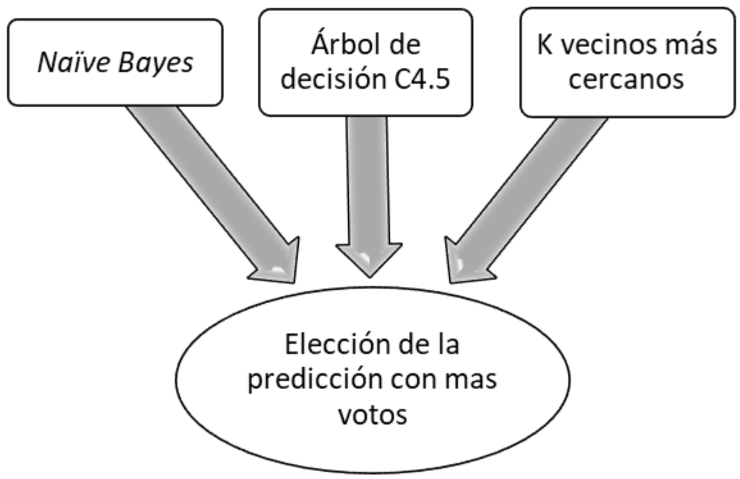
	Record 1	Record 2	…..	Record 250
Exercise 1	< 6	without delivering	…..	< 6
Exercise 2	≥ 6	≥ 6	…..	Without delivering
Exercise 3	< 6	≥ 6	…..	≥ 6
Exercise 4	≥ 6	without delivering	…..	≥ 6
Exercise 5	< 6	without delivering	…..	≥ 6
Final course evaluation	approved	failed	…..	approved

Source: Own elaboration

In the creation of the predictive models, combined machine learning techniques were used, specifically, the majority voting and stacking techniques.

The majority voting technique involves using predictions from underlying machine learning techniques and then selecting the prediction with the most votes, that is, the majority vote (García *et al.*, 2023). The underlying techniques used in this research are *Naive Bayes* (Sarmiento *et al.*, 2024), C4.5 decision trees (Timarán *et al*., 2023), and k nearest neighbors (Arengas *et al.*, 2024). The structure of the majority voting technique implemented in this article is shown in Figure 2.

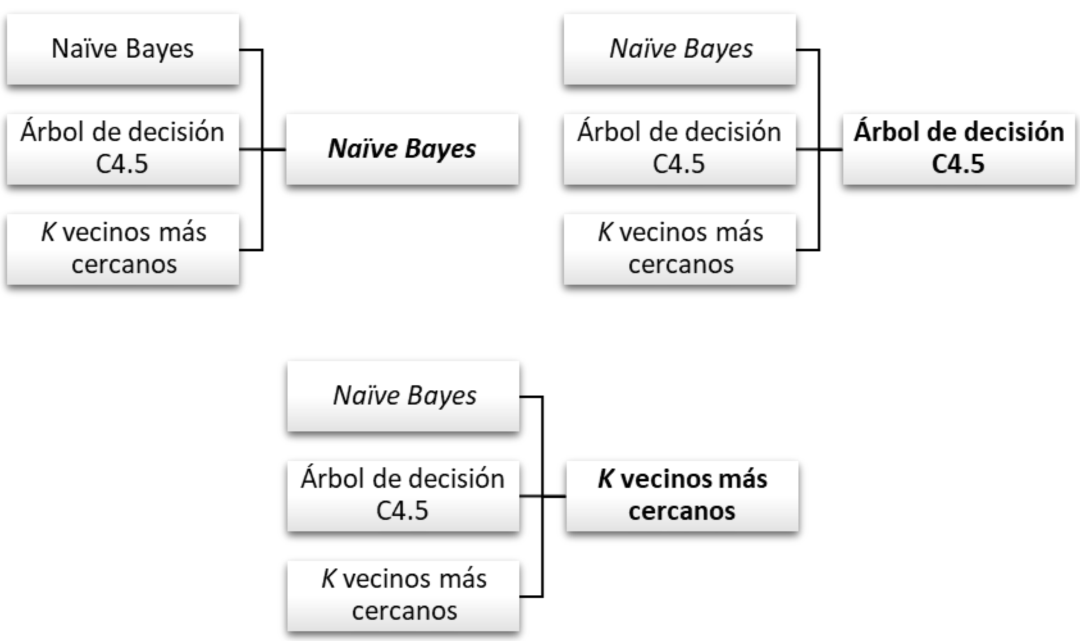
Figure 2. Structure of the majority voting technique in predictive models.



Source: Own elaboration

Stacking is a machine learning algorithm that uses predictions from base techniques (first phase) as inputs to another machine learning technique (second phase) (Cruz, 2024*).* In the first phase, the three machine learning techniques used in majority voting are employed, and in the second phase, the same three techniques from the first phase are used, resulting in three different stacking techniques, as shown in Figure 3.

Figure 3. Structure of stacking techniques in predictive models.



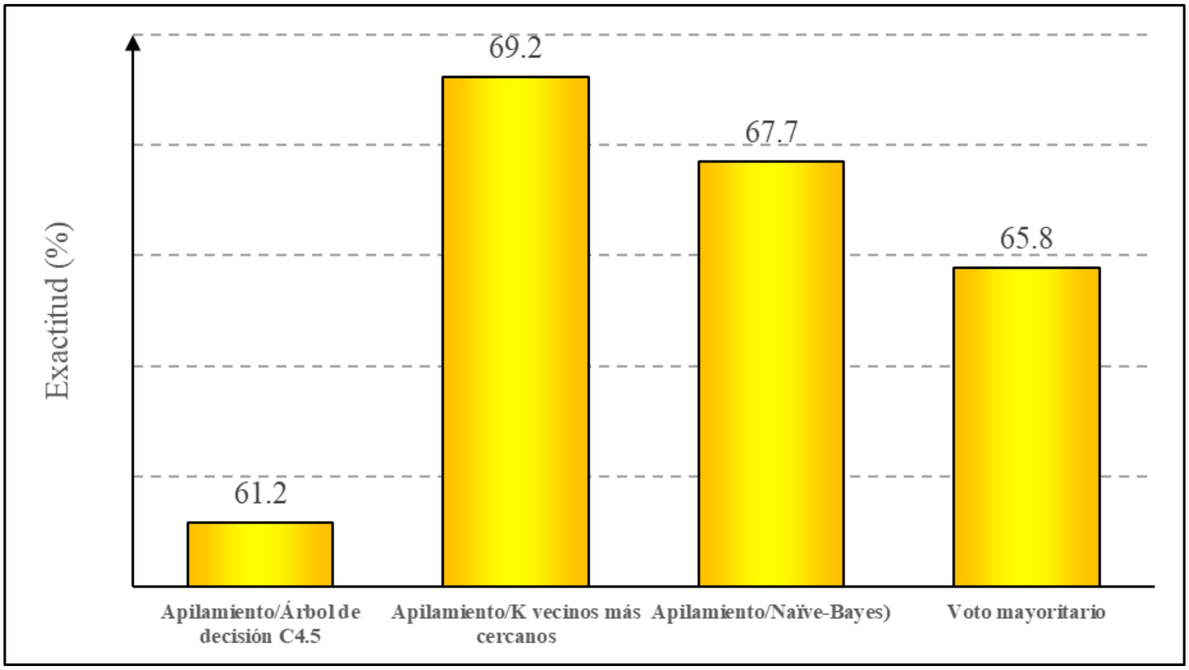
Source: Own elaboration

These models and all the analyses in this article were done with the support of the free software known as Weka (Nizar *et al.,* 2024). Furthermore, they require evaluation; that is, they need metrics to verify that the given predictions are acceptable. The metrics used in this research are accuracy, true positive rate, and true negative rate (Yajure , 2023; Daza *et al.,* 2024). Prediction accuracy is calculated by dividing the number of correct predictions by the total number of predictions made by the model. The true positive rate is the number of predictions of student passing divided by the total number of actual passing predictions. The true negative rate is the same ratio but considering the number of predictions of students failing. For each machine learning technique, these metrics are calculated using cross-validation with 10 groups (Sierra *et al*., 2024), similar to how it has been used in similar studies in the literature (Parraga , 2024). This method involves randomly dividing the data into 10 groups. Nine groups are used to build the model, and predictions are made for the remaining group. This process is repeated 10 times. The accuracy, true positive rate, and true negative rate are the averages obtained in the cross-validation process for each of the machine learning techniques used in this research. Specifically, for the *k*- nearest neighbors technique, which serves as the basis for the majority voting and stacking techniques, as well as for the technique used in the second phase of the stacking technique, the parameter *k* that maximizes the accuracy of the predictions in its respective cross-validation was selected. The results of these metrics with the machine learning techniques are presented in the following section.

Results

The results in this section employ predictive models developed using combined machine learning techniques, including majority voting and three stacking techniques described in the previous section. Figures 4, 5, and 6 present the accuracy, true positive rate, and true negative rate for the majority voting technique (Figure 2) and the three stacking techniques (Figure 3). For the stacking techniques, to distinguish which technique is used in the second phase, it is proposed to represent them as “stacking/technique in the second phase.” For example, “stacking/ *Naive Bayes* ” means that the stacking technique is used, with the *Naive Bayes* algorithm employed in the second phase.

Figure 4. Accuracy of predictive models of academic performance with combined machine learning techniques using cross-validation.

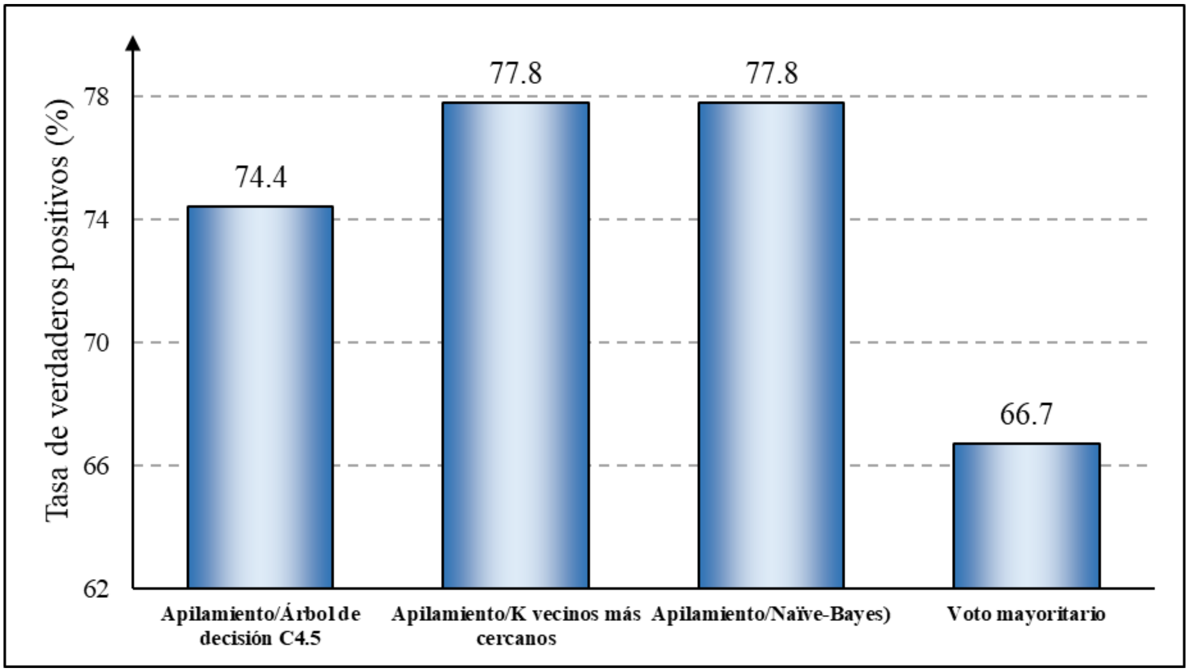


Source: Own elaboration

Figure 4 shows that the Stacking/ *K* nearest neighbors technique achieves the highest accuracy compared to the others, at 69.2%. It can also be seen that the Stacking/C4.5 decision tree technique is lower than the others (61.2%).

Revista Iberoamericana para la Investigación y el Desarrollo Educativo
ISSN 2007 - 7467

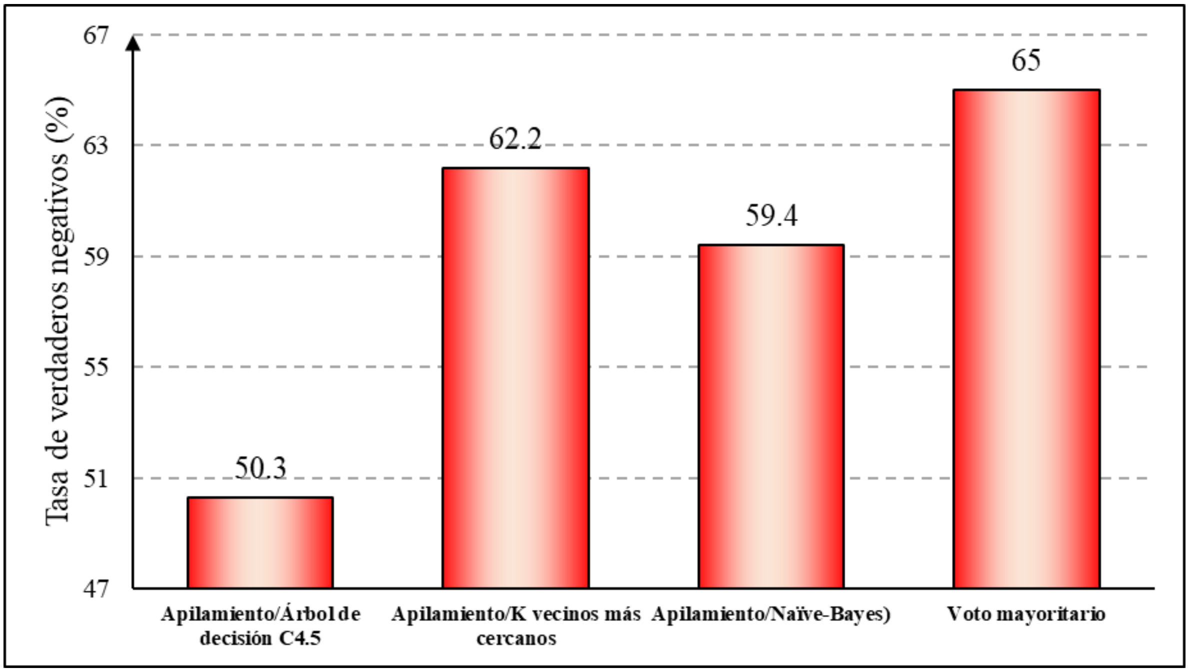
Figure 5. True positive rate of predictive models of academic performance with combined machine learning techniques using cross-validation.



Source: Own elaboration

The true positive rates with stacking techniques are the highest, as shown in Figure 5, specifically those using *Naive Bayes* and *k* nearest neighbors in the second phase. Furthermore, the true positive rate with the majority voting technique is significantly lower than the others.

Figure 6. True negative rate of predictive models of academic performance with combined machine learning techniques using cross-validation.

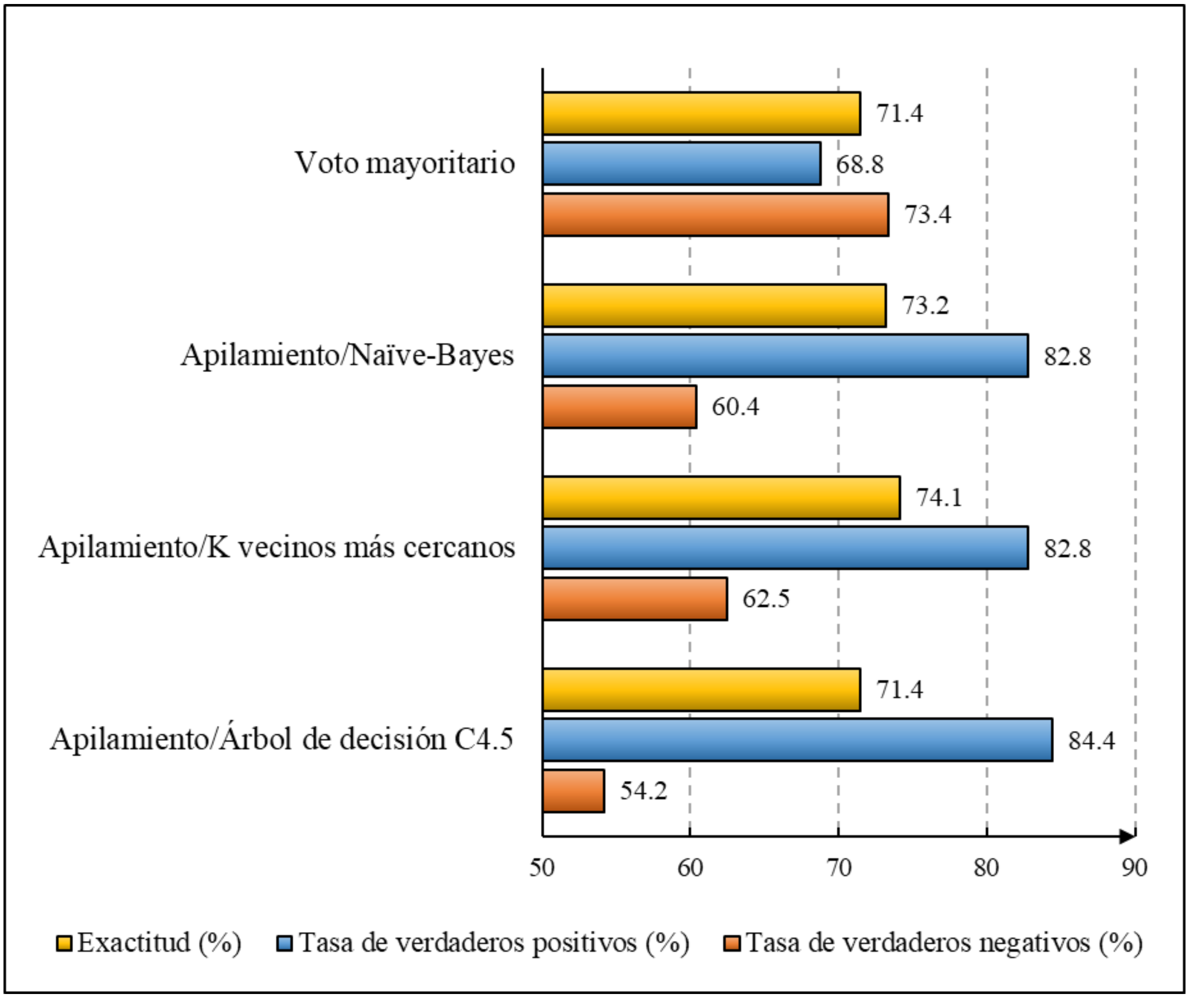


Source: Own elaboration

In Figure 6, the true negative rate with the majority voting technique (65%) is the highest compared to the other stacking techniques.

To demonstrate the usefulness of the models developed using combined machine learning techniques, predictions were made for 108 university students enrolled in the same course but from a later semester; that is, an independent sample. Upon completion, information regarding their passing grades was collected and compared with the model predictions. This yielded metrics for accuracy, true positive rate, and true negative rate, which are shown in Figure 7.

Figure 7. Metrics of predictive models of academic performance with combined machine learning techniques.



Source: Own elaboration

The figure above shows that the true positive rate tends to be higher in the models compared to the other metrics with values; similarly, the true negative rate tends to be lower. Furthermore, the accuracy of the predictions tends to be similar across all techniques, being higher with the stacking technique that uses k nearest neighbors in the second phase (around 70%).

Discussion

The results obtained through cross-validation show that the stacking technique using the *k*- nearest neighbors algorithm in the second phase has the highest accuracy (69.2%), while the techniques using *Naive Bayes* and *k*- nearest neighbors have the highest true positive rate (77.8%), and the majority voting technique has the highest true negative rate (65%). Therefore, with the data used, no single technique employed in this research achieves the best metric values in all cases. Instead, it is necessary to choose the most appropriate technique based on the desired metric value. In this regard, when predicting academic performance with another dataset (108 students) using the developed models, the highest accuracy is obtained with the stacking technique using the k-nearest neighbors method in the second phase, at 74.1%. The highest true positive rate (84.4%) was obtained with the stacking technique using the C4.5 decision tree. and the highest true negative rate was calculated to be 73.4% using the majority voting technique.

In this study, the exercises were developed during the first three weeks of the 18-week course. This means the predictions were made when the course was 17% complete, allowing for the early identification of students with a high probability of failing so that instructors can implement targeted pedagogical interventions. It is also important to note that collecting assessments of academic exercises is routine for instructors and doesn't require more complex data collection tools, making this type of methodology more accessible for building predictive models.

There are studies that use metrics to evaluate predictive models similar to those employed in this research. In the study conducted by Gil and Quintero (2023), they achieved a maximum accuracy of 59% in their models and a true positive rate with a maximum value of 71%. To achieve this, they used seven attributes or variables to develop the predictive models, and the collected data was divided into 70% for training and 30% for testing. Villarrasa *et al*. (2024) obtained maximum accuracy, true positive rate, and true negative rate values of 81.4%, 79%, and 83.33%, respectively. They used 14 variables and divided the data into 80% for training and 20% for testing. In contrast to these studies, the present study used cross-validation, adding randomization to the experiments and offering potentially more robust results for the metrics. Furthermore, the test data consisted of students predicting their course performance at the time of the study, rather than previously stored records. This research required five predictor variables and yielded metric values higher than those obtained by Gil and Quintero (2023). While the metrics obtained in Villarrasa *et al.* (2024)

have higher values than those obtained in the present study, and the number of variables used in this article is much lower. It should also be noted that previous studies use variables that are not as easily collected by teachers, unlike in this study, where exercise evaluations are routinely collected by teachers.

Finally, the methodology developed in this study could be used in other courses and educational modalities in order to detect students with academic difficulties so that the teacher can implement appropriate support strategies to improve their performance.

Conclusions

This study designed a methodology for developing models to predict the academic performance of Mexican students at the end of a course. The model employed in-class exercises and a combination of machine learning techniques, specifically majority voting and three stacking methods. To evaluate the models, the metrics of accuracy, true positive rate, and true negative rate were used. These metrics were initially obtained using cross-validation and subsequently by applying the models to a different dataset—a test dataset.

When the models were evaluated with cross validation, the highest values of accuracy (69.2%), true positive rate (77.8%) and true negative rate (65%) were obtained with the stacking technique that in the second phase has the *k* nearest neighbors algorithm, with the stacking techniques that have *Naïve Bayes* and *k* nearest neighbors, and with the majority voting technique, respectively.

When the models were evaluated with the test data, the highest values of accuracy (74.1%), true positive rate (84.4%) and true negative rate (73.4%) resulted with the stacking technique that has in the second phase the *k* nearest neighbors technique, with the C4.5 decision tree stacking technique and with the majority voting technique, respectively.

The information gathered from students was collected during the first three weeks of the course, representing 17% of the coursework. This facilitates the early identification of students who may be experiencing academic difficulties, allowing teachers to provide timely interventions to improve their performance. This type of methodology could be replicated in other courses and educational levels. Furthermore, it allows for understanding the specific needs of students in order to personalize tutoring programs or other support strategies depending on the subject. Predictive models can contribute to a more efficient use of educational resources, enabling a focus on the students who need them most.

It should be noted that, despite the progress shown in this research, there are several areas for improvement regarding the topic addressed in this study. These include employing the methodology with other types of machine learning techniques or with different student characteristics in order to improve the evaluation metrics.

Future lines of research

It should be noted that, despite the progress achieved in this article, the scope was limited by the institutional context, sample size, and the type and number of variables. However, further research in this area is possible. Increasing the amount of data used in each model is considered pertinent. Furthermore, it may be beneficial to increase the number of machine learning techniques employed in the majority voting and stacking methods. Additionally, other types of attributes (academic, social, behavioral, etc.) that may influence student academic performance could be incorporated.

References

Arengas, J. M., López, M. y Guzmán, R. (2024). Impacto del preprocesamiento en la clasificación automática de textos usando aprendizaje supervisado y reuters 21578. *Revista colombiana de tecnologías avanzadas (RCTA)*, *1*(43), 110–118. https://doi.org/10.24054/rcta.v1i43.2506

Contreras, L. E., Fuentes, H. J. y Rivas, E. (2021). Análisis del rendimiento académico mediante técnicas de aprendizaje automático con métodos de ensamble. *Revista Boletín Redipe, 10*(13), 171-190. https://doi.org/10.36260/rbr.v10i13.1737

Cruz, L. D. (2024). Ensamble híbrido de métodos para la clasificación morfotaxonómica explicable de macroinvertebrados de agua dulce. *Encuentro Internacional De Educación En Ingeniería*. https://doi.org/10.26507/paper.3565

Daza, J., Castro, J. E. y Ávila, H. (2024). Optimizando el aprendizaje de los lenguajes de programación. Un enfoque basado en la analítica de datos para los estudiantes de Ingeniería de Sistemas en la Fundación Universitaria Los Libertadores. *Perspectivas, 9*(24), 234-256. https://doi.org/10.26620/uniminuto.perspectivas.9.24.2024.234-256

Del Carpio, R., (2024). Predicción del rendimiento académico utilizando modelos de aprendizaje automático: Una revisión sistemática de la literatura. 593 Digital Publisher CEIT, 9(6), 1038-1054, https://dialnet.unirioja.es/servlet/articulo?codigo=9842504

García, F. A., Escobar, J. L., Gallegos, C. M. y Hernández, E. S. (2023). El enfoque de aprendizaje conjunto en la detección de fallas en cajas de engranajes. *Revista Universidad y Sociedad*, *15*(3), 325-333. http://scielo.sld.cu/pdf/rus/v15n3/2218-3620-rus-15-03-325.pdf

Gil, V. D. y Quintero, C. (2023). Análisis de variables asociadas al rendimiento académico en cursos universitarios virtuales. *Formación universitaria*, *16*(4), 33-42. https://dx.doi.org/10.4067/s0718-50062023000400033

Nizar, N., Zainudin, A. D., Albada, A. y Shan, C. M. (2024). Forecasting Short-Term FTSE Bursa Malaysia Using WEKA. *Information Management and Business Review*, *16*(2), 104-114. https://doi.org/10.22610/imbr.v16i2(I)S.3773

Parraga, D. (2024). Modelos Predictivos de Rendimiento Académico Universitario Mediante Aprendizaje Automático. *Vitalia Revista Científica y Académica*, *5*(2), 974–991. https://doi.org/10.61368/r.s.d.h.v5i2.204

Salas, R. A., Salas, E. P., Salas, R. D. y Vargas Y. D. (2019). Análisis de la Aplicación Web Para la Estimación Puntual por medio de la Ciencia de Datos. *Revista Dilemas Contemporáneos: Educación, Política y Valores*, *6*(2), Artículo 33. https://dilemascontemporaneoseducacionpoliticayvalores.com/index.php/dilemas/article/view/705

Sarmiento, J. D., Tibaduiza, D. A., Anaya, M., Gómez, J. D., Pérez, M. P., Sanchez, C. E. y Eslava, J. S. (2024). Uso de Machine Learning en la identificación y clasificación de microplásticos enfocado en el poliestireno expandido. En Ingeniería: una transición hacia el futuro. Asociación Colombiana de Facultades de Ingeniería – ACOFI. https://doi.org/10.26507/paper.4046

Sierra, M. A., Quintana, K. P., Hernández, J. A., Enríquez, L. B., Pérez, M. D. y Arzate, C. (2024). Validación de un modelo de inteligencia artificial para la predicción de la mortalidad del paciente con sepsis. *Medicina Interna de México*, *40*(3), 171-178. https://doi.org/10.24245/mim.v40i3.9023

Timarán, R., Chaves, A. y Ordoñez-Erazo, H. (2023). Decision Tree Algorithm Moderately Coupled to PostgreSQL DBMS. *Revista Facultad De Ingeniería*, *32*(66), e16777. https://revistas.uptc.edu.co/index.php/ingenieria/article/view/16777

Vargas, A. A. y Prieto, J. C. (2024). Predicción del rendimiento académico estudiantil usando algoritmos de aprendizaje supervisado en una universidad de la selva peruana. *Revista*

Amazonía Digital, *3*(1), e292. https://revistas.unamad.edu.pe/index.php/rad/article/view/292

Villarrasa, I., García, X., Liébana, E. y Monfort, G. (2024). Predicción del rendimiento académico en educación secundaria mediante el análisis de árboles de decisión. Educación XX1, 27(1), 253-279. https://doi.org/10.5944/educxx1.33351

Yajure, C. A. (2023). Selección del modelo óptimo de predicción de la relación de desempeño de una planta solar fotovoltaica. Un enfoque multicriterio basado en algoritmos de aprendizaje automático. *Ciencia, Ingeniería y Aplicaciones*, *6*(2), 7-29. https://doi.org/10.22206/cyap.2023.v6i2.2935

Zambrano, J. M., Arroyo, M. V., Jalil, N. J. y Castro, I. E. (2024). Perfiles de educación universitaria: Un análisis comparativo. Revista De Ciencias Sociales, 30(2), 291-304. https://dialnet.unirioja.es/servlet/articulo?codigo=9603967

Contribution Role	Author(s)
Conceptualization	Andrés Rico Páez
Methodology	Andrés Rico Páez, Nora Diana Gaytán Ramírez (same)
Software	NOT APPLICABLE
Validation	Andrés Rico Páez, Nora Diana Gaytán Ramírez (same)
Formal Analysis	Andrés Rico Páez, Nora Diana Gaytán Ramírez (same)
Investigation	Andrés Rico Páez (principal), Nora Diana Gaytán Ramírez (supporting)
Resources	Andrés Rico Páez.
Data curation	Andrés Rico Páez (principal), Nora Diana Gaytán Ramírez (supporting)
Writing - Preparing the original draft	Andrés Rico Páez (principal), Nora Diana Gaytán Ramírez (supporting)
Writing - Reviewing and Editing	Andrés Rico Páez, Nora Diana Gaytán Ramírez (same)
Display	Andrés Rico Páez (principal), Nora Diana Gaytán Ramírez (supporting)
Supervision	Andrés Rico Páez
Project Management	Andrés Rico Páez (principal), Nora Diana Gaytán Ramírez (supporting)
Acquisition of funds	Andrés Rico Páez