

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

Artículos científicos

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 prever 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.

Fecha Recepción: Septiembre 2025

Fecha Aceptación: Febrero 2026

Introducción

En años recientes se han dado grandes avances en la tecnología como los dispositivos móviles, internet y redes de telefonía móvil, lo cual impacta en diversas áreas. Una de las áreas que ha recibido una gran influencia debido al desarrollo tecnológico es la educación. El crecimiento en las tecnologías de la información y la comunicación ha facilitado las actividades de aprendizaje en la adquisición de conocimientos y además, en la recopilación de registros de usuarios. En instituciones educativas, las herramientas digitales han permitido mejorar los métodos de acceso y almacenamiento de información de estudiantes tales como calificaciones, asistencia, créditos cursados entre muchos otros atributos (Salas *et al.*, 2019). El procesamiento y análisis de información permite entender los datos para dar posibles



soluciones a problemas en el área educativa que pueden influir en los procesos de enseñanza y aprendizaje. Dichos análisis permiten identificar ciertas similitudes, diferencias y patrones en los datos que pueden ayudar al desarrollo de políticas educativas y a mejorar las prácticas pedagógicas (Zambrano *et al.*, 2024). De entre los análisis de registros educativos empleados en la mejora de los procesos de enseñanza se encuentra la predicción del rendimiento académico de estudiantes mediante las técnicas de aprendizaje automático (Del Carpio, 2024).

La predicción del rendimiento académico de estudiantes universitarios ha aumentado debido al interés de investigadores e instituciones educativas en mejorar la estrategia pedagógica e identificar tempranamente a estudiantes en peligro de reprobación que requieren algún tipo de apoyo o intervención académica (Zambrano *et al.*, 2024). Para realizar la predicción del rendimiento académico se han empleado técnicas de aprendizaje automático (Vargas y Prieto, 2024). Estas técnicas permiten construir modelos de predicción a partir de un conjunto de registros llamados datos de entrenamiento con el objetivo de predecir y validar con otro conjunto de datos conocido como datos de prueba. Existen trabajos en esta línea de investigación, como el realizado en (Gil y Quintero, 2023), en el cual los autores utilizan datos de estudiantes que participaron en cuatro cursos universitarios virtuales. Los autores construyeron modelos de predicción con diferentes técnicas de aprendizaje automático a partir de siete variables académicas predictoras. Villarrasa *et al.* (2024) realizaron modelos de predicción del rendimiento académico realizando una técnica de aprendizaje automático conocida como árbol de decisión empleando 14 variables predictoras.

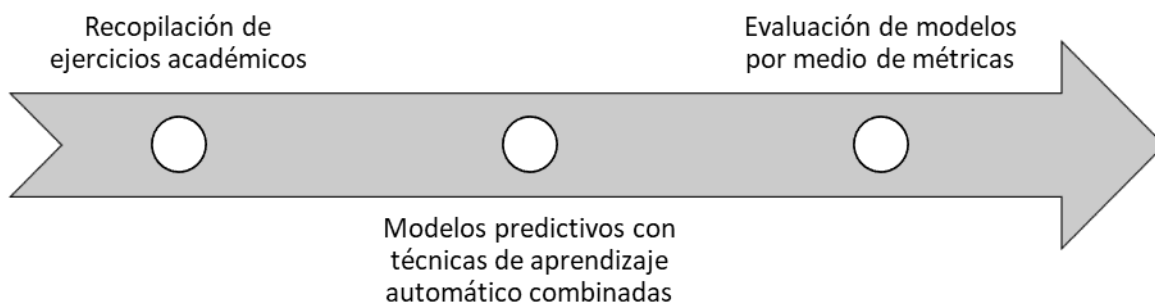
Actualmente, existen trabajos que emplean modelos predictivos del rendimiento académico desarrollados con técnicas de aprendizaje automático combinadas (Contreras *et al.*, 2021), es decir, combinan técnicas como si fuera una sola con el propósito de mejorar las predicciones. En México existe poco avance en el desarrollo de modelos predictivos con este tipo de técnicas debido, entre otras causas, a la falta de conocimiento de dichas metodologías. Esto ha ocasionado un cierto retraso en comparación con otros países en cuanto al procesamiento de datos académicos con el potencial beneficio de mejora de la enseñanza. Por lo anterior, en esta investigación se plantean las siguientes preguntas de investigación: ¿Cómo elaborar modelos predictivos del rendimiento académico de estudiantes en México utilizando ejercicios realizados en clase y empleando técnicas de aprendizaje automático combinadas? ¿De qué manera evaluar los modelos predictivos desarrollados? De esta

manera, el objetivo de este artículo es elaborar una metodología para construir modelos predictivos del rendimiento académico de estudiantes de México al finalizar un curso mediante ejercicios realizados en clase con técnicas de aprendizaje automático combinadas y proponer métricas para evaluar dichos modelos.

Metodología

La metodología empleada en la presente investigación consiste en recabar información de 250 estudiantes de una universidad pública en México acerca de sus evaluaciones obtenidas en cinco ejercicios académicos de una asignatura de Matemáticas en las primeras semanas. Con esta información se realizan modelos de predicción de la aprobación o reprobación de la asignatura, es decir, de su desempeño académico al final del curso. Este tipo de herramientas facilita a los profesores identificar a los estudiantes con mayor probabilidad de que reprueben el curso, y en consecuencia, realizar intervenciones que permitan regularizar a estos estudiantes en etapas tempranas del curso. En la Figura 1 se presenta de manera esquemática la metodología utilizada.

Figura 1. Esquema de la metodología empleada en la investigación.



Fuente: Elaboración propia

Los ejercicios se realizaron en las tres primeras semanas del curso, cuya duración es de 18 semanas. Se informó a los participantes que los datos recabados se utilizarían únicamente con fines de investigación educativa. Además, para proteger la confidencialidad y por razones éticas no se utilizaron sus nombres, sino que se les asoció un número identificador ya que las técnicas de aprendizaje automático no los requieren.

En este estudio, las evaluaciones de los ejercicios tuvieron un rango de calificación de 0 a 10. Se clasificaron como aquellas que son aprobatorias, es decir, son igual o más de seis (≥ 6), las que fueron reprobatorias o menor a seis (< 6) y los ejercicios que no fueron entregados por los estudiantes para su evaluación (sin entregar). Finalmente, la evaluación final del curso de cada estudiante tiene los valores de aprobado o reprobado. Parte de la estructura de los datos recabados de los estudiantes se presenta en la tabla 1.

Tabla 1. Estructura de datos recabados en la investigación.

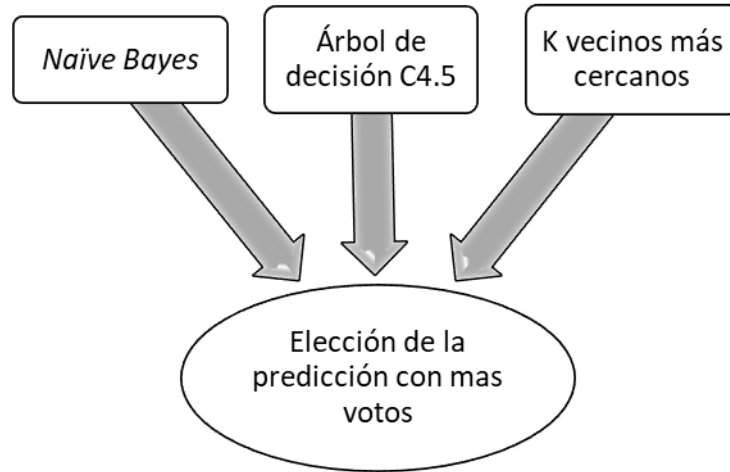
	Registro 1	Registro 2	Registro 250
Ejercicio 1	< 6	sin entregar	< 6
Ejercicio 2	≥ 6	≥ 6	Sin entregar
Ejercicio 3	< 6	≥ 6	≥ 6
Ejercicio 4	≥ 6	sin entregar	≥ 6
Ejercicio 5	< 6	sin entregar	≥ 6
Evaluación final del curso	aprobado	reprobado	aprobado

Fuente: Elaboración propia

En la realización de los modelos predictivos se emplearon técnicas de aprendizaje automático combinadas, específicamente, las técnicas de voto mayoritario y de apilamiento.

La técnica de voto mayoritario consiste en utilizar las predicciones de técnicas base de aprendizaje automático, posteriormente, elige la predicción que tenga la mayor cantidad de votos, es decir, la de voto mayoritario (García *et al.*, 2023). Las técnicas base empleadas para esta investigación son *Naïve Bayes* (Sarmiento *et al.*, 2024), árbol de decisión C4.5 (Timarán *et al.*, 2023) y *k* vecinos más cercanos (Arenegas *et al.*, 2024). La estructura de la técnica voto mayoritario que se realizó en este artículo se representa en la Figura 2.

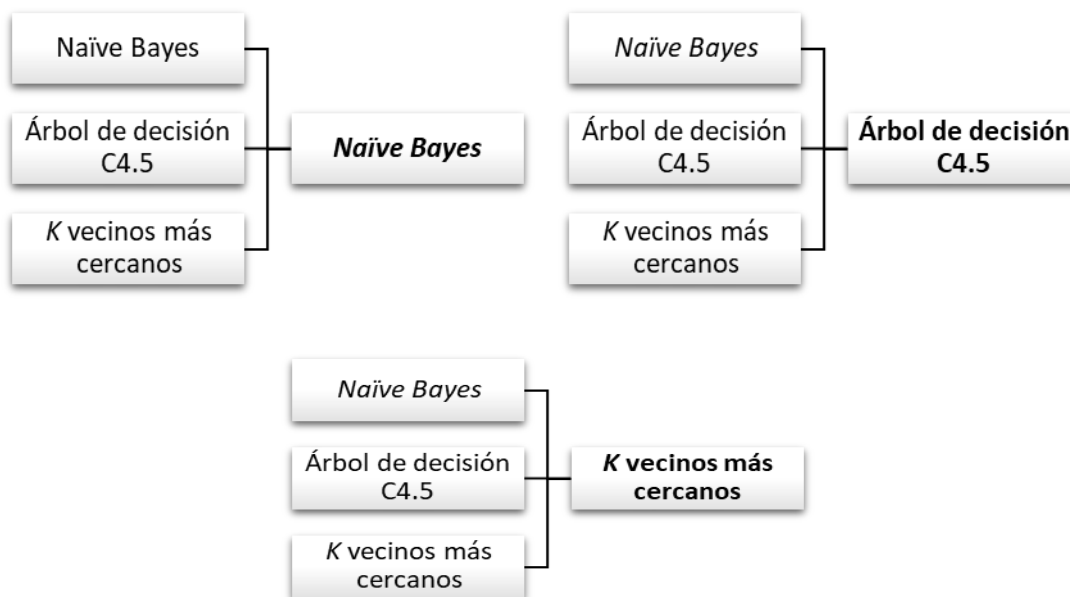
Figura 2. Estructura de la técnica voto mayoritario en los modelos predictivos.



Fuente: Elaboración propia

La técnica de apilamiento o *stacking* es un algoritmo de aprendizaje automático que emplea las predicciones de técnicas base (primera fase) como entradas a otra técnica de aprendizaje automático (segunda fase) (Cruz, 2024). En la primera fase se utilizan las tres técnicas de aprendizaje automático que las empleadas en voto mayoritario, y en la segunda fase, se utilizan las mismas tres técnicas de la primera fase obteniendo tres técnicas de apilamiento diferentes como se muestra en la Figura 3.

Figura 3. Estructura de las técnicas de apilamiento en los modelos predictivos.



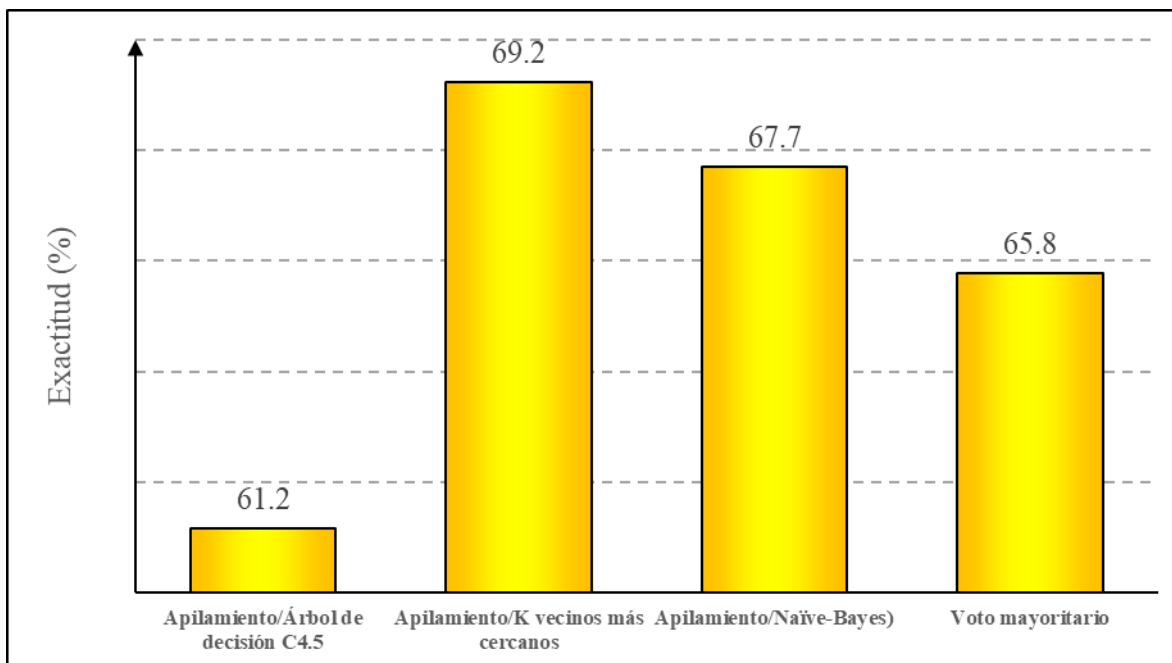
Fuente: Elaboración propia

Estos modelos y todos los análisis en este artículo se hicieron con el apoyo del software libre conocido como Weka (Nizar *et al.*, 2024). Asimismo, requieren ser evaluados, es decir, necesitan de métricas para verificar que las predicciones dadas sean aceptables. Las métricas que se emplean en esta investigación son la exactitud, la tasa de verdaderos positivos y la tasa de verdaderos negativos (Yajure, 2023; Daza *et al.*, 2024). La exactitud de las predicciones es aquella que contabiliza las predicciones que fueron acertadas dividido entre el total de las predicciones realizadas por el modelo. La tasa de verdaderos positivos son las predicciones de aprobación del estudiante dividido entre todas las predicciones de aprobación reales. La tasa de verdaderos negativos es la misma relación pero considerando las predicciones de reprobación del estudiante. Para cada técnica de aprendizaje automático, estas métricas se calculan utilizando la validación cruzada considerando 10 grupos (Sierra *et al.*, 2024) de manera parecida a como ha sido empleada en trabajos similares en la literatura (Parraga, 2024). Este método consiste en dividir los datos aleatoriamente en 10 grupos, se utilizan 9 grupos para construir el modelo y se hacen predicciones del grupo restante, este proceso se repite en 10 ocasiones. La exactitud, la tasa de verdaderos positivos y la tasa de verdaderos negativos son el promedio de las obtenidas en el proceso de validación cruzada para cada una de las técnicas de aprendizaje automático usadas en esta investigación. En el caso específico de la técnica de k vecinos más cercanos que se utiliza como técnica base para la técnica de voto mayoritario y las de apilamiento, así como la que se emplea en la segunda fase de la técnica de apilamiento, se seleccionó el parámetro k que maximice la exactitud de las predicciones en su respectiva validación cruzada. En la siguiente sección se muestran los resultados de estas métricas con las técnicas de aprendizaje automático.

Resultados

En los resultados de esta sección se emplean los modelos predictivos realizados con técnicas de aprendizaje automático combinadas que incluyen voto mayoritario y tres técnicas de apilamiento descritas en la sección anterior. En las Figuras 4, 5 y 6 se presentan la exactitud, la tasa de verdaderos positivos y la tasa de verdaderos negativos con la técnica de voto mayoritario (Figura 2) y las tres técnicas de apilamiento (Figura 3). En el caso de las técnicas de apilamiento, para distinguir la técnica de que se utiliza en la segunda fase, se propone representar como “apilamiento/técnica en la segunda fase”, por ejemplo, “apilamiento / *Naïve Bayes*” significa que se utiliza la técnica de apilamiento teniendo en la segunda fase el algoritmo *Naïve Bayes*.

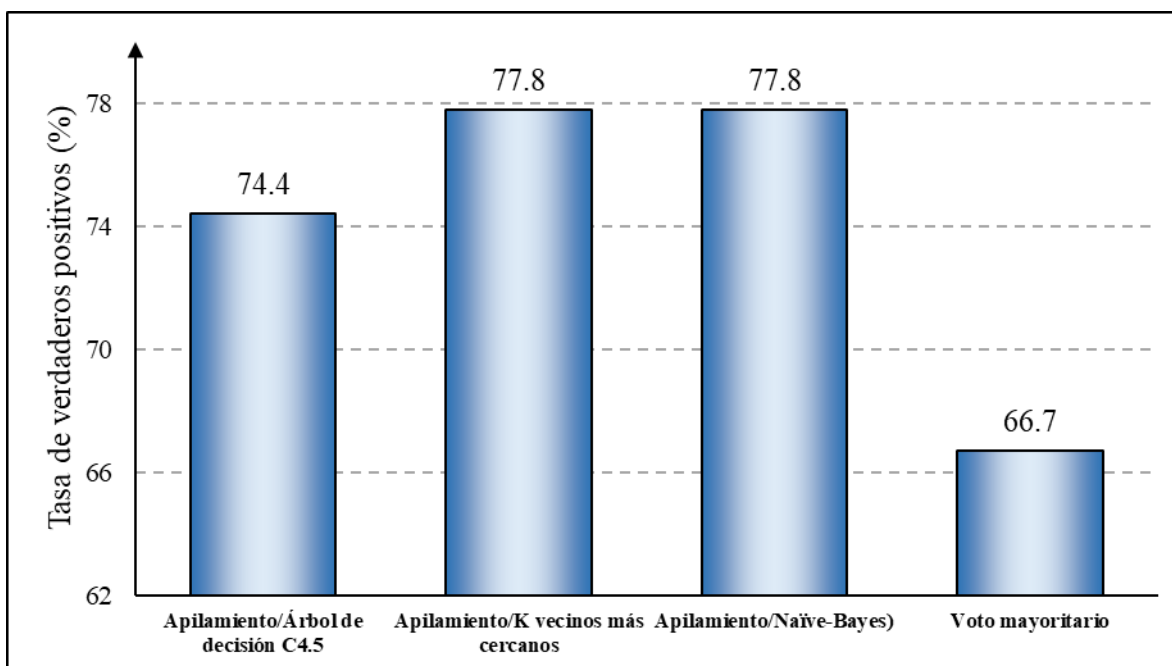
Figura 4. Exactitud de los modelos predictivos del rendimiento académico con las técnicas de aprendizaje automático combinadas utilizando validación cruzada.



Fuente: Elaboración propia

En la Figura 4 se observa que la técnica de Apilamiento/ K vecinos más cercanos es la que consigue la exactitud mayor en comparación con las demás, la cual es de 69.2%. También se puede apreciar que la técnica Apilamiento/Árbol de decisión C4.5 es inferior a las demás (61.2%).

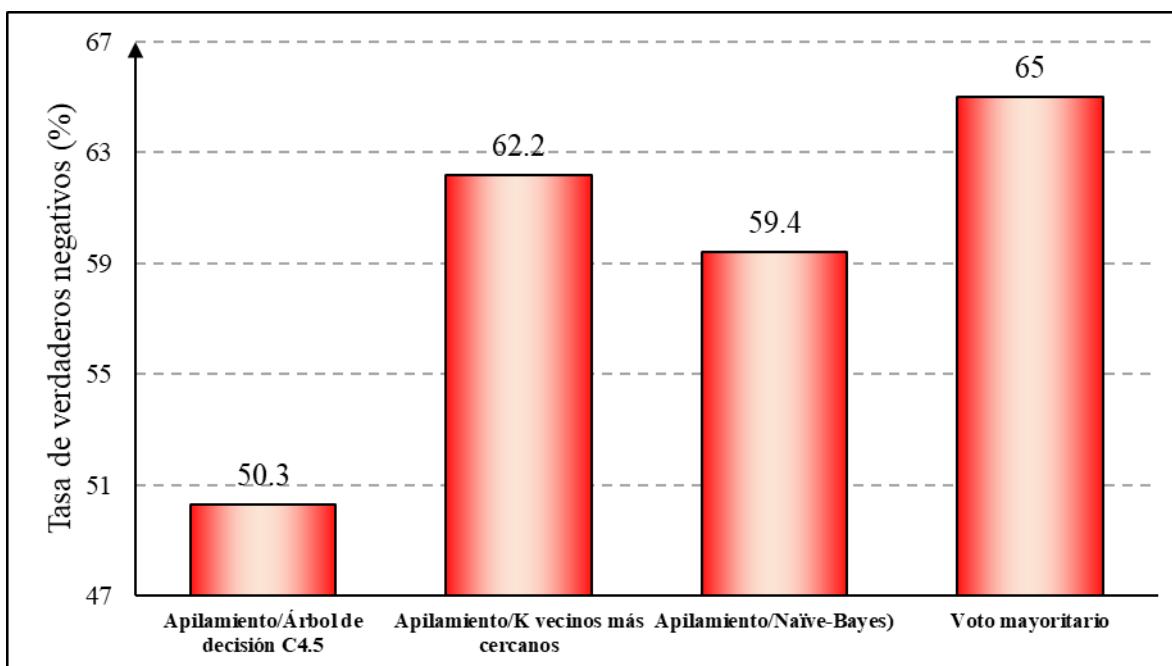
Figura 5. Tasa de verdaderos positivos de los modelos predictivos del rendimiento académico con las técnicas de aprendizaje automático combinadas utilizando validación cruzada.



Fuente: Elaboración propia

Las tasas de verdaderos positivos con las técnicas de apilamiento son las más altas como se muestra en la Figura 5, específicamente las que utilizan *Naïve Bayes* y *k* vecinos más cercanos en la segunda fase. Además, se observa que la tasa de verdaderos positivos con la técnica de voto mayoritario es mucho menor que las demás.

Figura 6. Tasa de verdaderos negativos de los modelos predictivos del rendimiento académico con las técnicas de aprendizaje automático combinadas utilizando validación cruzada.

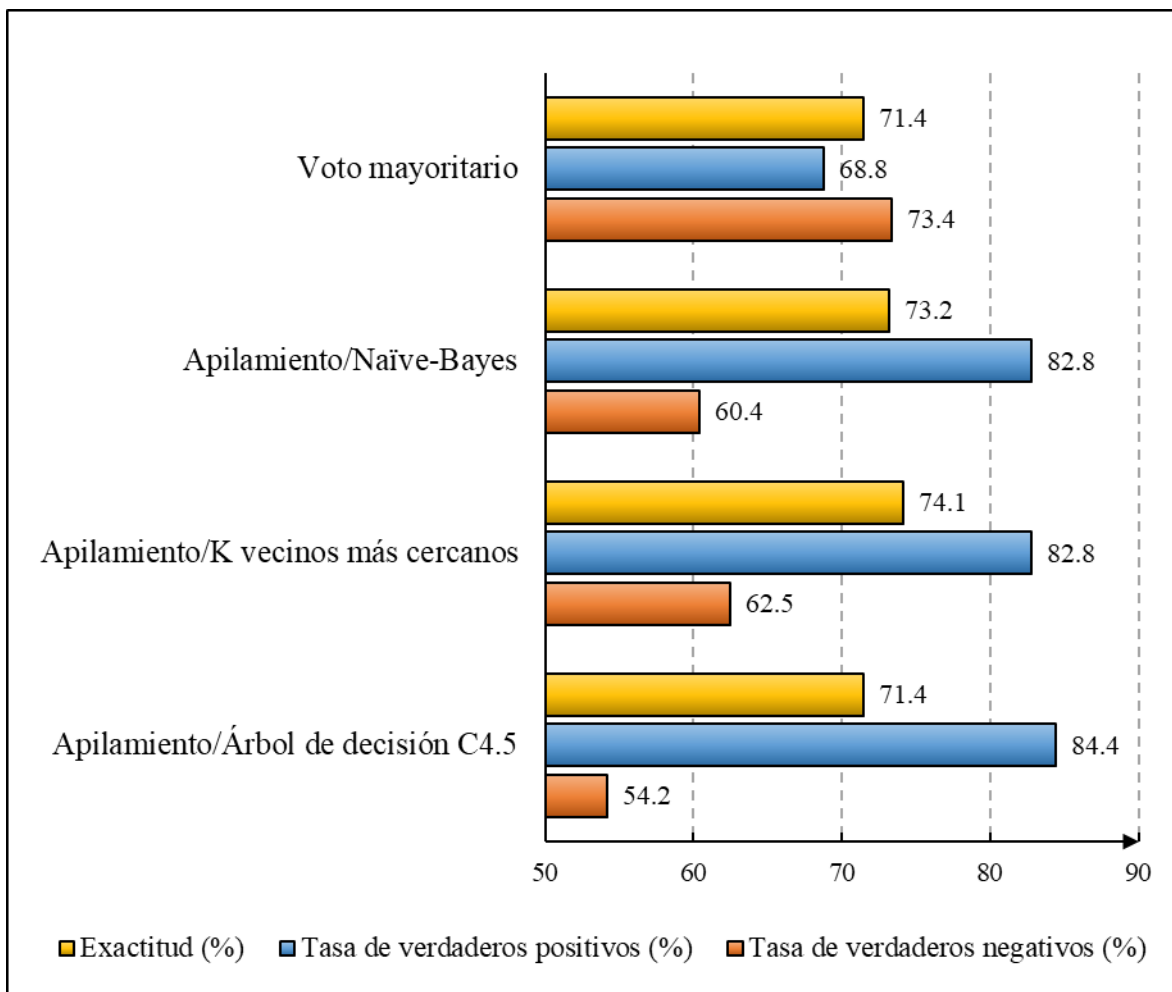


Fuente: Elaboración propia

En la Figura 6, la tasa de verdaderos negativos con la técnica de voto mayoritario (65%) es la mayor en comparación a las demás técnicas de apilamiento.

Para mostrar la utilidad de los modelos desarrollados con las técnicas de aprendizaje automático combinadas se realiza la predicción de 108 estudiantes universitarios del mismo curso pero de un periodo de impartición posterior, es decir, es una muestra independiente. Al terminar, se recopiló la información referente a su aprobación para comparar con las predicciones de los modelos con lo cual se obtuvieron las métricas de exactitud, tasa de verdaderos positivos y tasa de verdaderos negativos, las cuales se muestran en la Figura 7.

Figura 7. Métricas de los modelos predictivos del rendimiento académico con las técnicas de aprendizaje automático combinadas.



Fuente: Elaboración propia

En la figura anterior se observa que la tasa de verdaderos positivos tiende a ser mayor en los modelos en comparación con las demás métricas con valores, de forma análoga, la tasa de verdaderos negativos muestra una tendencia a ser menor. Además, la exactitud de las predicciones tiende a ser similar con todas las técnicas, siendo mayor con la técnica de apilamiento que utiliza k vecinos más cercanos en la segunda fase (alrededor de 70%).

Discusión

Los resultados obtenidos con validación cruzada muestran que la técnica de apilamiento que tiene en la segunda fase el algoritmo k vecinos tiene una mayor exactitud (69.2%), las que tienen *Naïve Bayes* y k vecinos más cercanos tienen la más alta tasa de verdaderos positivos (77.8%) y la técnica de voto mayoritario es la que tiene el valor mayor de tasa de verdaderos negativos (65%). De esta manera, con los datos utilizados, no se tiene una técnica, de las empleadas en esta investigación, que obtenga los mejores valores de las métricas en todos los casos, sino que es necesario elegir la técnica más adecuada de acuerdo con la métrica que se quiera con mayor valor. En este sentido, cuando se predice el rendimiento académico con otro grupo de datos (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 tasa de verdaderos positivos más alta se obtuvo de 84.4% con la técnica de apilamiento con árbol de decisión C4.5. y la tasa de verdaderos negativos mayor se calculó de 73.4% con la técnica de voto mayoritario.

En la investigación realizada los ejercicios se elaboraron en las primeras tres semanas de un total de 18 semanas que dura el curso, es decir, las predicciones se realizaron cuando el curso había avanzado un 17%, lo que permite identificar de manera temprana los estudiantes que tienen altas probabilidades de reprobar el curso para que los profesores puedan realizar intervenciones pedagógicas específicas. También, es importante tener en cuenta que la recopilación de las evaluaciones de los ejercicios académicos es habitual en los profesores sin requerir utilizar otras herramientas más complejas de recopilación de información lo que hace más accesible utilizar este tipo de metodologías para construir modelos predictivos.

Existen trabajos que utilizan métricas para evaluar los modelos predictivos similares a las empleadas en la presente investigación. En el estudio realizado por Gil y Quintero (2023) consiguieron una exactitud máxima en sus modelos de 59% y una tasa de verdaderos positivos con un valor máximo de 71%. Para ello, emplearon siete atributos o variables para realizar los modelos de predicción y los datos recabados los partieron en 70% para entrenamiento y 30% para prueba. Villarrasa *et al.* (2024) obtuvieron los valores máximos de exactitud, tasa de verdaderos positivos y tasa de verdaderos negativos de 81.4%, 79% y 83.33% respectivamente. Emplearon 14 variables y dividieron los datos en 80% de entrenamiento y 20% para prueba. En contraste con estos trabajos, en el presente estudio se

utilizó la validación cruzada agregando aleatoriedad a los experimentos ofreciendo resultados potencialmente más robustos de las métricas. Además, los datos de prueba fueron estudiantes a predecir su rendimiento en el curso al momento de realizar el estudio en lugar de registros almacenados con anterioridad. En esta investigación se requirieron cinco variables predictoras y se obtuvieron valores de las métricas superiores a los obtenidos por Gil y Quintero (2023). Si bien las métricas obtenidas en Villarrasa *et al.* (2024) son mayores a los obtenidos en la presente investigación, el número de variables empleadas en el presente artículo es mucho menor. También, se debe notar que los trabajos anteriores emplean variables que su recopilación no es tan fácil de obtener por los profesores, a diferencia de este trabajo, donde las evaluaciones de los ejercicios son recopiladas por los profesores de manera habitual.

Por último, la metodología desarrollada en este estudio podría emplearse en otros cursos y modalidades educativas con el propósito de detectar estudiantes con dificultades académicas y el profesor pueda implementar estrategias de apoyo adecuadas para mejorar su rendimiento.

Conclusiones

En este trabajo se diseñó una metodología en la que se desarrollaron modelos de predicción del desempeño académico al terminar un curso de estudiantes de México empleando ejercicios hechos en clase y técnicas de aprendizaje automático combinadas. Éstas fueron voto mayoritario y tres técnicas de apilamiento. Para evaluar los modelos se utilizaron las métricas de exactitud, tasa de verdaderos positivos y tasa de verdaderos negativos. Estas métricas se obtuvieron, primeramente, con el método de validación cruzada, y posteriormente, aplicando los modelos realizados con otro conjunto de datos, esto es, con datos de prueba.

Cuando se evaluaron los modelos con validación cruzada los valores más altos de exactitud (69.2%), tasa de verdaderos positivos (77.8%) y tasa de verdaderos negativos (65%) se obtuvieron con la técnica de apilamiento que en la segunda fase tiene el algoritmo k vecinos más cercanos, con las de apilamiento que tienen *Naïve Bayes* y k vecinos más cercanos, y con la de voto mayoritario, respectivamente.

Cuando se evaluaron los modelos con los datos de prueba los valores más altos de exactitud (74.1%), tasa de verdaderos positivos (84.4%) y tasa de verdaderos negativos (73.4%) resultaron con la técnica de apilamiento que tiene en la segunda fase la técnica k

vecinos más cercanos, con la técnica de apilamiento con árbol de decisión C4.5 y con la técnica de voto mayoritario, respectivamente.

La información recabada de los estudiantes se llevó a cabo en las primeras tres semanas del curso, equivalente a un 17% de avance en la asignatura. Esto facilita detectar de manera temprana a los estudiantes que puedan tener dificultades escolares y que los profesores brinden intervenciones oportunas para mejorar su rendimiento. Este tipo de metodologías podría ser replicada en otros cursos y niveles educativos. Además, permite entender las necesidades específicas de estudiantes para personalizar programas de asesorías o de otro tipo de estrategias de apoyo dependiendo de la asignatura. Los modelos predictivos pueden contribuir a hacer uso más eficiente de recursos de la institución educativa lo que permite enfocar en estudiantes que más lo requieran.

Se debe notar que, a pesar de los avances mostrados en la presente investigación, existen varias líneas de mejora sobre la temática tratada en este estudio. Entre las cuales se incluyen emplear la metodología con otro tipo de técnicas de aprendizaje automático o con otras características de estudiantes con el fin de mejorar las métricas de evaluación.

Futuras líneas de investigación

Cabe mencionar que, a pesar de los progresos conseguidos en el presente artículo, se limitó al contexto institucional, al tamaño de muestra, al tipo y cantidad de variables. No obstante, es posible continuar profundizando sobre esta línea de investigación. Se considera pertinente incrementar la cantidad de datos empleados en cada modelo. Además, puede ser de interés, incrementar el número de técnicas de aprendizaje automático que se emplean en la técnica de voto mayoritario y en la de apilamiento. Asimismo, se puede emplear otro tipo de atributos (académicos, sociales, conductuales, etc.) que puedan influir en el desempeño académico del estudiante.

Referencias

- 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 morfotaxonomica 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](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://dilemascontemporaneoseduccionpoliticayvalores.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>

Rol de Contribución	Autor (es)
Conceptualización	Andrés Rico Páez
Metodología	Andrés Rico Páez, Nora Diana Gaytán Ramírez (igual)
Software	NO APLICA
Validación	Andrés Rico Páez, Nora Diana Gaytán Ramírez (igual)
Análisis Formal	Andrés Rico Páez, Nora Diana Gaytán Ramírez (igual)
Investigación	Andrés Rico Páez (principal), Nora Diana Gaytán Ramírez (apoya)
Recursos	Andrés Rico Páez.
Curación de datos	Andrés Rico Páez (principal), Nora Diana Gaytán Ramírez (apoya)
Escritura - Preparación del borrador original	Andrés Rico Páez (principal), Nora Diana Gaytán Ramírez (apoya)
Escritura - Revisión y edición	Andrés Rico Páez, Nora Diana Gaytán Ramírez (igual)
Visualización	Andrés Rico Páez (principal), Nora Diana Gaytán Ramírez (apoya)
Supervisión	Andrés Rico Páez
Administración de Proyectos	Andrés Rico Páez (principal), Nora Diana Gaytán Ramírez (apoya)
Adquisición de fondos	Andrés Rico Páez