

<https://doi.org/10.23913/ride.v12i24.1196>

*Artículos científicos*

## **Modelos predictivos progresivos del rendimiento académico de estudiantes universitarios**

***Progressive Predictive Models of the Academic Performance of University Students***

***Modelos predictivos progresivos de desempenho acadêmico de estudantes universitários***

**Andrés Rico Páez**

Instituto Politécnico Nacional, México

[aricop.ipn@gmail.com](mailto:aricop.ipn@gmail.com)

<https://orcid.org/0000-0002-6450-318X>

### **Resumen**

El objetivo de esta investigación fue desarrollar modelos predictivos progresivos del rendimiento académico de estudiantes universitarios de México y evaluarlos para distintas técnicas de aprendizaje automático. En este estudio se recopiló calificaciones de actividades académicas de 260 estudiantes universitarios para crear modelos de predicción de los resultados académicos mediante técnicas de aprendizaje automático. Se construyeron los modelos en diferentes etapas a lo largo del curso y se evaluaron empleando la exactitud en la predicción de 112 estudiantes de un curso posterior. Se observó una exactitud de hasta 70.5 % en un tiempo de 21 % del total de la duración del curso. Este tipo de metodología puede ser replicada para diferentes tipos de cursos debido a que el registro de calificaciones es común en casi todos ellos. Además, esta metodología es flexible en cuanto a la elección de la etapa temporal en la cual realizar las predicciones, sin perder el compromiso con la exactitud. Así, se puede efectuar en etapas tempranas para detectar problemas con el rendimiento académico y evitar, en la medida de lo posible, la reprobación y deserción de estudiantes.

**Palabras clave:** aprendizaje automático, modelo matemático, prevención, rendimiento académico.

## Abstract

The objective of this research was to develop progressive predictive models of the academic performance of university students in Mexico and evaluate them for different machine learning techniques. In this study, grades of academic activities of 260 university students were collected to create prediction models of academic results using machine learning techniques. The models were built at different stages throughout the course and were evaluated using the accuracy of the predictions by applying it to the prediction of 112 students in a subsequent course. An accuracy of up to 70.5 % was observed in a time of 21 % of the total duration of the course. This type of methodology can be replicated for different types of courses because the recording of grades is common in almost all courses. In addition, this methodology is flexible in terms of choosing the time stage in which to make the predictions, maintaining a compromise between the accuracy of the predictions and that they be made at the earliest possible stage to detect problems with academic performance, avoiding, in as far as possible, the failure and desertion of students.

**Keywords:** machine learning, mathematical model, prevention, academic performance.

## Resumo

O objetivo desta pesquisa foi desenvolver modelos preditivos progressivos do desempenho acadêmico de estudantes universitários no México e avaliá-los para diferentes técnicas de aprendizado de máquina. Neste estudo, foram coletadas notas de atividades acadêmicas de 260 estudantes universitários para criar modelos de previsão de resultados acadêmicos usando técnicas de aprendizado de máquina. Os modelos foram construídos em diferentes etapas ao longo do curso e testados usando a precisão de previsão de 112 alunos de um curso subsequente. Foi observada acurácia de até 70,5% em um tempo de 21% da duração total do curso. Esse tipo de metodologia pode ser replicada para diferentes tipos de cursos, pois o registro de notas é comum a quase todos eles. Além disso, essa metodologia é flexível quanto à escolha do momento de realização das previsões, sem perder o compromisso com a

precisão. Assim, pode ser feito precocemente para detectar problemas com o desempenho acadêmico e evitar, na medida do possível, a reprovação e a evasão dos alunos.

**Palavras-chave:** aprendizaje de máquina, modelo matemático, prevenção, desempenho acadêmico.

**Fecha Recepción:** Noviembre 2021

**Fecha Aceptación:** Mayo 2022

---

## Introducción

Actualmente, el desarrollo tecnológico ha provocado que, en el área educativa, surjan grandes cantidades de datos referentes a estudiantes, profesores y demás integrantes del proceso educativo. Comúnmente, estos datos son generados para ciertos objetivos y no se analizan, principalmente, porque se desconoce cómo hacerlo. Sin embargo, puede existir información potencialmente útil que puede beneficiar el proceso educativo en la disminución de la deserción estudiantil y mejora del rendimiento académico. La deserción estudiantil es un problema que está asociado a muchas variables y que perjudica a todos los actores del proceso educativo (Rivera, 2021). El rendimiento académico es la suma de distintos y complejos factores del entorno estudiantil (Garbanzo, 2007) y, en las instituciones educativas, es uno de los principales indicadores acerca de la calidad educativa. De esta manera, para las instituciones educativas es importante la recolección de datos para analizarlos y encontrar información que pueda mejorar su sistema educativo (Bakhshinategh, *et al.*, 2018). En años recientes, diversos trabajos han utilizado el análisis de datos para predecir el rendimiento escolar (Kalaivani, Priyadharshini y Selva, 2017; La Red *et al.*, 2015).

La predicción del rendimiento académico puede realizarse a distintos niveles de detalle, por ejemplo, para predecir resultados de tareas, exámenes o un curso completo (Asif *et al.*, 2017). En cualquier caso, es deseable la predicción del rendimiento académico porque permite identificar tempranamente a los estudiantes en riesgo de reprobar y permite realizar algún tipo de intervención que evite que abandonen sus estudios e incentivar su retención en la escuela.

Las técnicas de aprendizaje automático son aquellas que aprenden un modelo a partir de un conjunto de datos, y actualmente, se están utilizando para construir modelos de predicción de los resultados del aprendizaje de estudiantes (Xing *et al.*, 2015). Es decir, estas técnicas permiten construir modelos que aprenden a partir de datos provenientes de

ambientes educativos para luego predecir resultados de datos nuevos (Contreras, Fuentes y Rodríguez, 2020). Altujjar *et al.* (2016) construyeron un modelo predictivo utilizando la técnica de aprendizaje automático conocida como *algoritmo de árbol de decisión ID3* para predecir el bajo rendimiento en estudiantes universitarios. Hussain *et al.* (2018) realizaron predicciones de resultados de aprendizaje con distintos algoritmos de redes bayesianas y árboles de decisión a partir de datos socioeconómicos y demográficos de estudiantes universitarios. Usman *et al.* (2019) emplearon técnicas de árbol de decisión, *naive Bayes* y *k* vecinos más cercanos para predecir el rendimiento académico en función de su interacción con una plataforma educativa en Internet. Contreras *et al.* (2020) implementaron técnicas de aprendizaje automático como la *k* vecinos más cercanos para predecir el rendimiento de estudiantes de ingeniería industrial.

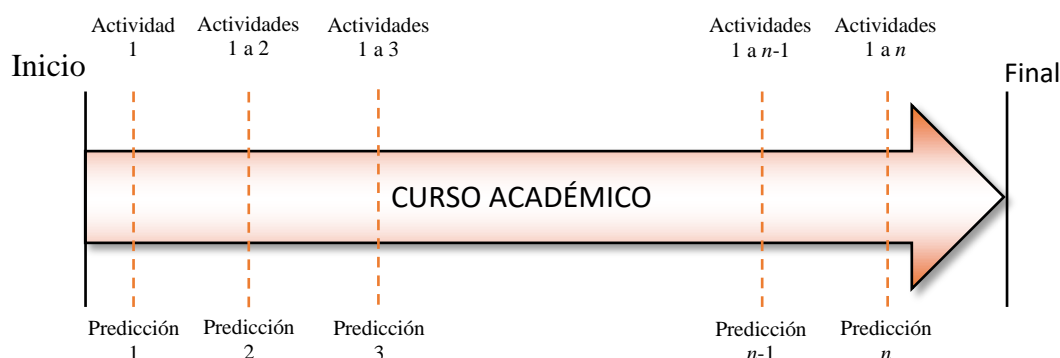
En la literatura existen varios estudios acerca de la predicción del rendimiento académico que han aportado información útil para la comprensión y planificación de los procesos educativos. No obstante, las bajas calificaciones de estudiantes siguen siendo un problema en las instituciones educativas, principalmente en las universidades. Esto se puede reflejar en la gran cantidad de estudiantes universitarios que reprueban y se dan de baja, lo cual es más frecuente en los primeros años de estudio (Silva, 2011). De esta manera, existe la necesidad de desarrollar metodologías para la predicción de los resultados del aprendizaje que permitan el análisis y uso de la información escolar para mejorar la calidad educativa. Sin embargo, en nuestro país, existen pocos trabajos que hacen uso de técnicas de aprendizaje automático en la construcción de modelos predictivos del rendimiento académico (Ayala, López y Menéndez, 2021; Juárez, Cortés y Coronilla, 2014; Valero, Vargas y García, 2010).

En esta investigación se plantean las siguientes preguntas: ¿cómo realizar modelos de predicción progresivos del desempeño académico de estudiantes de una universidad de México? y ¿cómo evaluar los modelos realizados mediante las técnicas de aprendizaje automático propuestas? De esta manera, el objetivo de la presente investigación es desarrollar modelos de predicción progresivos del desempeño académico de estudiantes de una universidad mexicana y evaluarlos para distintas técnicas de aprendizaje automático.

## Metodología

En este trabajo se propone una metodología que consiste en utilizar las calificaciones obtenidas en actividades académicas por estudiantes durante un curso de una universidad pública de México. La metodología consiste en realizar una predicción 1 con un modelo predictivo construido a partir de la actividad 1, después se realiza una predicción 2 con un modelo predictivo realizado con las actividades desde la 1 hasta la 2, este proceso se repite hasta utilizar las actividades 1 hasta la  $n$ . Esta metodología se ilustra en la figura 1. Por lo tanto, mientras el curso progresa se adicionan más datos a los modelos predictivos.

**Figura 1.** Metodología propuesta para construir modelos predictivos progresivos del rendimiento académico en un curso



Fuente: Elaboración propia

En este estudio participaron 260 estudiantes de una universidad pública de México y se utilizaron 14 actividades académicas realizadas durante un curso. Estas actividades son realizadas en espacios de tiempo similares a lo largo del curso. Cada actividad se considera aprobada (A) si su calificación está entre 6 y 10; en caso contrario, se considera reprobada (R), y también se considera el caso en el que el estudiante no presente la actividad (NP). De esta manera, se construye una tabla con 260 registros y 15 columnas (atributos) que servirá para construir los modelos predictivos. Una muestra de esta se presenta en la tabla 1. Cada actividad se representa como “act” y el número de la actividad. Además, la aprobación del estudiante se representa con atributo “aprueba”, que puede tener los valores de “Sí” o “No”.

**Tabla 1.** Muestra de datos para los modelos predictivos del rendimiento académico

act1	act2	act3	act4	act5	act6	act7	act8	act9	act10	act11	act12	act13	act14	aprueba
R	A	R	A	A	R	A	A	NP	R	NP	A	R	R	Sí
NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	No
R	A	R	A	A	R	A	A	A	R	A	R	A	R	No
NP	NP	R	A	A	R	A	R	A	R	A	R	R	R	Sí
R	NP	R	A	A	R	A	R	A	A	A	A	R	R	Sí
NP	NP	A	NP	A	NP	NP	R	A	R	A	A	R	R	Sí
R	R	R	NP	A	R	A	A	A	A	A	R	R	R	Sí
NP	NP	NP	NP	NP	R	NP	NP	NP	NP	NP	NP	NP	NP	No
R	NP	A	A	NP	R	A	NP	NP	R	A	NP	NP	NP	No
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:

Fuente: Elaboración propia

## Resultados

Un modelo predictivo se construye utilizando un conjunto de datos conocidos como *datos de entrenamiento* y una técnica de aprendizaje automático. En este estudio, los datos de entrenamiento son la tabla con 260 registros y las técnicas de aprendizaje automático que se utilizan son *naïve Bayes*, *k* vecinos más cercanos y árbol de decisión C4.5 (Hernández, Ramírez y Ferri, 2004). Todos los análisis de datos mostrados en este artículo se realizaron con el apoyo del *software* libre Weka (Witten, Frank y Hall, 2005).

En esta investigación se recabaron 14 actividades académicas de estudiantes, por lo que se construyeron 14 grupos de datos por medio de la tabla 1. El primer grupo consiste en los datos de la primera columna de la tabla de 260 registros (actividad 1) y de la columna del atributo “aprueba”. El segundo grupo contiene los datos de las dos primeras columnas de la tabla (actividades 1 a la 2) y de la columna del atributo “aprueba”. Así sucesivamente, hasta que se incluyen todas las columnas de la tabla (14 actividades y el atributo “aprueba”). Con cada uno de los 14 grupos de datos de entrenamiento se construye un modelo predictivo para cada una de las técnicas de aprendizaje automático. De esta manera, se va construyendo un modelo predictivo, primeramente, con la actividad 1, después, con las actividades 1 y 2, y así sucesivamente, hasta utilizar las 14 actividades académicas. Es decir, los modelos predictivos van incrementando progresivamente actividades académicas en sus datos de entrenamiento.

En esta investigación se utiliza la exactitud de las predicciones como métrica de evaluación del desempeño de los modelos predictivos, la cual se define como el número de predicciones que fueron correctas dividido entre las predicciones totales (Durairaj y Vijitha, 2014).

La técnica  $k$  vecinos más cercanos utiliza el parámetro  $k$ , una forma de seleccionarlo es eligiendo aquel que consiga un mayor valor en la exactitud de las predicciones. Para esto se emplea la validación cruzada, la cual consiste en particionar, de manera aleatoria, los datos en una cantidad fija de particiones; se reserva una partición para realizar las predicciones y las restantes para construir el modelo predictivo, esta acción se repite dejando una partición diferente para realizar las predicciones. La exactitud se calcula promediando las exactitudes obtenidas con cada partición. Para estos experimentos se utilizó una validación cruzada con 10 particiones, ya que ha sido empleada en trabajos similares (Márquez *et al.*, 2012; Mueen, Zafar y Manzoor, 2016). Cada grupo de datos tiene 260 registros, así que para cada uno de los 14 grupos se calcula la exactitud para diversos valores de  $k$  (1, 2, 3..., 260) y se elige, para cada grupo, el valor de  $k$  en donde se obtiene la mayor exactitud, tal y como se muestra en la tabla 2.

**Tabla 2.** Valores de  $k$  donde se obtiene la mayor exactitud con la validación cruzada mediante la técnica  $k$  vecinos más cercanos

Cantidad de actividades académicas	Valor de $k$ donde se obtuvo la mayor exactitud
Actividad 1	1
Actividades 1 a la 2	3
Actividades 1 a la 3	11
Actividades 1 a la 4	68
Actividades 1 a la 5	90
Actividades 1 a la 6	74
Actividades 1 a la 7	36
Actividades 1 a la 8	18
Actividades 1 a la 9	87
Actividades 1 a la 10	6
Actividades 1 a la 11	155
Actividades 1 a la 12	164
Actividades 1 a la 13	115
Actividades 1 a la 14	105

Fuente: Elaboración propia

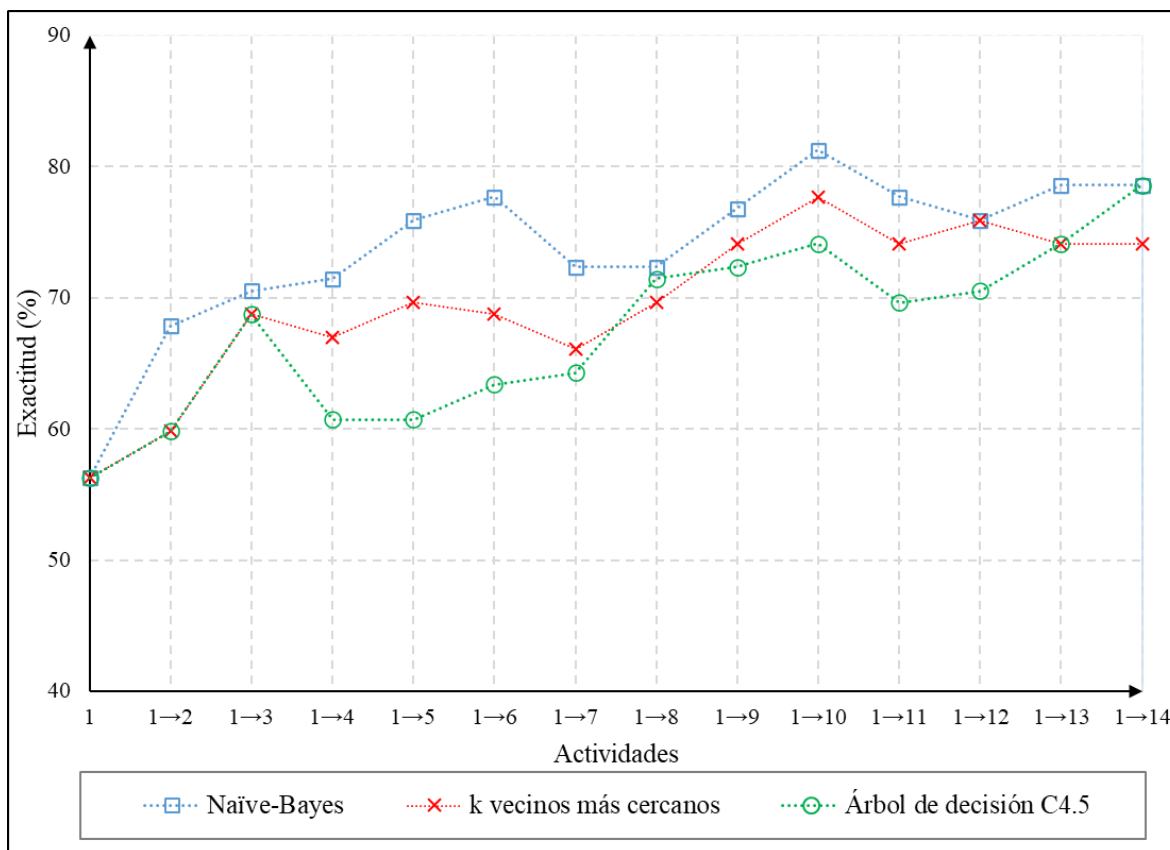
Con los cálculos anteriores, se aplican las tres técnicas de aprendizaje automático a los 14 grupos de datos de entrenamiento para construir modelos predictivos del rendimiento académico. Estos modelos se emplearon para predecir el rendimiento académico de 112 estudiantes universitarios que toman el mismo curso con la misma cantidad de actividades, pero de un semestre posterior al semestre en donde se recopilaban los datos de entrenamiento. Este tipo de datos se conocen como *datos de prueba*, es decir, son datos diferentes a los de entrenamiento y a los cuales se les harán las predicciones.

Después de realizar las predicciones, se recopilaron los resultados al final del curso obtenidos por los 112 estudiantes y se contabilizaron las predicciones que fueron correctas. De esta manera, se calculó la exactitud en los datos de prueba a partir de los modelos predictivos construidos con cada uno de los 14 grupos de datos de entrenamiento y con cada una de las tres técnicas de aprendizaje automático. Se debe notar que los modelos predictivos para cada uno de los 14 grupos representan modelos predictivos progresivos del rendimiento



académico porque van cambiando conforme se aumenta el número de actividades a lo largo del curso. En la figura 2 se presenta la exactitud de los modelos predictivos progresivos con diferentes técnicas de aprendizaje automático. En el eje de las abscisas se indica la cantidad de actividades académicas utilizadas para construir los modelos predictivos, por ejemplo, para cuando se empleó de la actividad 1 a la 3 se representó con la notación 1→3.

**Figura 2.** Exactitud de los modelos predictivos progresivos del rendimiento

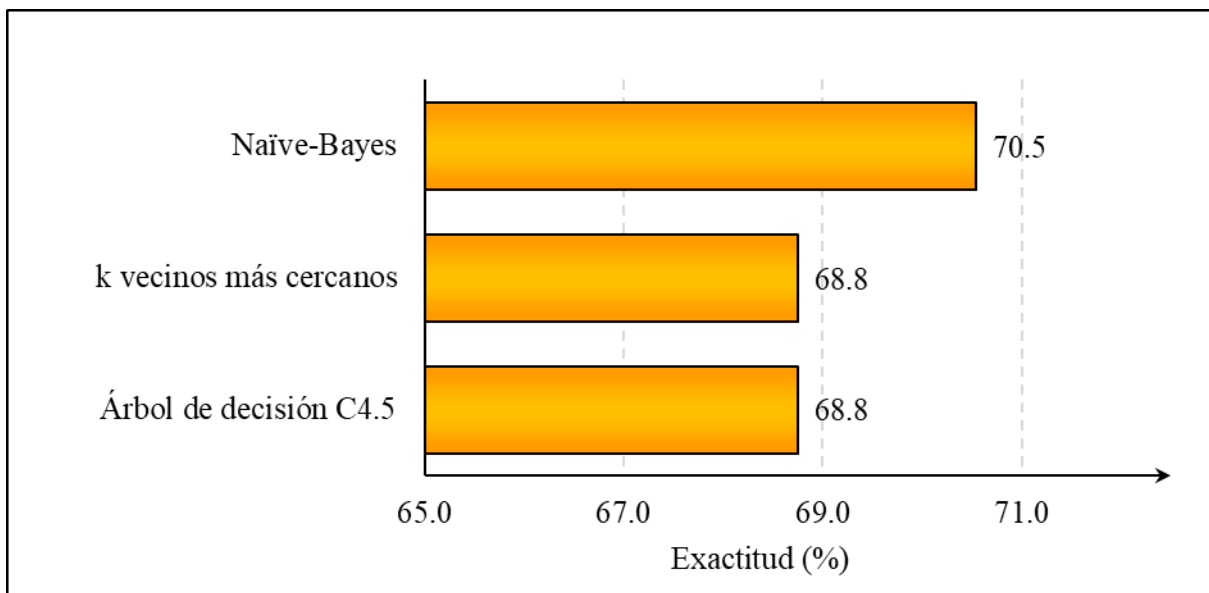


Fuente: Elaboración propia

En la figura 2 se observa que existe una cierta tendencia a que, entre mayor sea la cantidad de actividades que tengan los modelos predictivos, mayor será la exactitud de las predicciones, sin importar qué técnica de aprendizaje automático se utilice. Es decir, los modelos predictivos tienden a tener una mayor exactitud en las predicciones conforme va progresando el curso y se va añadiendo más información. También, se puede notar que los modelos predictivos progresivos del rendimiento académico realizados con la técnica *naive* Bayes tienen un mejor desempeño en cuanto a la exactitud de las predicciones. El valor de exactitud mayor es de 81.25 % y se consigue con la técnica *naive* Bayes y cuando se utilizan las actividades de la 1 a la 10. No obstante, la realización de la actividad 10 está cerca del

final del curso, así que la predicción de un estudiante cerca del final de un curso limita las acciones que se pueden realizar para evitar su reprobación. De esta manera, es necesario mantener un compromiso entre la identificación temprana del rendimiento académico del estudiante y la exactitud de las predicciones. Un valor de interés es cuando se utilizan las actividades 1 a la 3 para crear modelos predictivos debido a que con las tres técnicas generan un valor de exactitud de alrededor de 70 % (figura 3) y se consigue con las primeras tres actividades. Es decir, del inicio del curso hasta la tercera actividad ha pasado alrededor de 21 % de tiempo de la duración del curso ( $100\% \times 3/14$ ), lo que permite la identificación temprana de estudiantes en peligro de reprobación.

**Figura 3.** Exactitud de los modelos predictivos del rendimiento académico con las primeras tres actividades académicas



Fuente: Elaboración propia

## Discusión

En los resultados obtenidos se puede observar cómo la exactitud de los modelos predictivos tiene una tendencia creciente conforme se aumenta el número de actividades, debido a que se va adicionando más información a los modelos conforme progresa el curso, y aunque en algunas ocasiones la exactitud se mantiene o disminuye, de manera global tiende a aumentar. Los modelos predictivos progresivos realizados con la técnica *naive* Bayes tienen un mejor desempeño en cuanto a la exactitud de las predicciones debido a que en la mayoría

de los casos la exactitud es superior a las otras dos técnicas. Esto concuerda con lo observado por Osmanbegović y Suljić (2012) y Mueen *et al.* (2016), quienes, con cantidades de datos similares, obtuvieron valores mayores de exactitud con esta técnica en comparación con otras.

En los modelos predictivos progresivos no se requiere llegar al final del curso o incluso hasta las últimas actividades académicas para obtener predicciones aceptables para identificar a posibles estudiantes que reprobarán el curso. En este sentido, se observa que tan solo con las primeras tres actividades académicas se elaboran modelos predictivos de hasta 70.5 % con la técnica *naive* Bayes. Esto es importante porque no solo permite detectar estudiantes en peligro de reprobación, sino que también le brinda a las instituciones educativas y profesores un periodo de tiempo razonable para realizar las intervenciones necesarias sobre estudiantes específicos.

En la literatura se han observado estudios que realizan predicciones del desempeño académico utilizando técnicas de aprendizaje automático con actividades académicas. Sharma y Vishwakarma (2017) realizaron un estudio en donde participaron 70 estudiantes. Dividieron los datos en 50 registros para construir el modelo predictivo y 20 registros para predecir el rendimiento académico. Así, obtuvieron una exactitud de 90 % empleando todas las actividades académicas hasta la mitad del curso, es decir, este valor de exactitud se obtuvo a 50 % de la duración del curso. Del Campo *et al.* (2017) utilizaron 124 registros de estudiantes y obtuvieron, mediante validación cruzada, una exactitud máxima de 71.57 % a 50 % de tiempo de haber iniciado el curso. Alcaraz *et al.* (2020) utilizaron 78 registros de estudiantes y, con validación cruzada, obtuvieron una exactitud máxima de 79.5 % a 25 % de tiempo de haber iniciado el curso. En estos trabajos se utilizaron solamente los datos de entrenamiento para calcular la exactitud. A diferencia de estos trabajos, en esta investigación se utilizaron los modelos en datos de prueba, es decir, de estudiantes de un curso posterior al curso de donde se recabaron los datos de entrenamiento. Cabe mencionar que, si bien la exactitud de las predicciones en estos trabajos fue mayor a la obtenida en el presente artículo (70 %), en esta investigación se consiguió dicha exactitud en 21 % de tiempo de haber iniciado el curso, es decir, se obtuvo esta exactitud en una etapa más temprana que en los otros trabajos. También, en esta investigación se muestra el comportamiento de la exactitud de las predicciones en diferentes etapas del curso mediante la representación de actividades

académicas, lo que permite seleccionar un número de actividades que mantenga un compromiso entre un valor de exactitud alto y una realización temprana de las predicciones.

En este estudio, las causas que llevaron a reprobar a los estudiantes pudieron haber sido un mal diseño de las actividades académicas, la actuación del profesor, el estilo de aprendizaje, entre otros, debido a que el fracaso escolar depende de muchos factores (Antelm, Cacheiro y Gil, 2015). Por lo que es recomendable que el profesor realice actividades que faciliten a los estudiantes la adquisición de conocimientos y habilidades que les permitan construir su proceso de aprendizaje para mejorar su rendimiento académico.

## Conclusiones

En esta investigación se planteó una metodología que emplea actividades académicas durante un curso para realizar predicciones del rendimiento académico de estudiantes. Para realizar estas predicciones, se mostró cómo desarrollar modelos de predicción progresivos del desempeño académico de estudiantes de una universidad de México. Estos modelos se construyeron utilizando 260 registros de estudiantes y empleando las técnicas de aprendizaje automático *naive* Bayes, *k* vecinos más cercanos y árbol de decisión C4.5. Para evaluar estos modelos se utilizó la exactitud de las predicciones obtenida al aplicar dichos modelos en la predicción del desempeño académico de 112 estudiantes. La exactitud se obtuvo para diferentes cantidades de actividades académicas durante el curso.

Se observó que la exactitud tuvo una tendencia a incrementarse conforme avanzaba el curso, y que utilizando las primeras tres actividades se puede obtener una exactitud de hasta 70.5 % en 21 % de tiempo de haber iniciado el curso y se hizo en una etapa más temprana que en otros artículos similares revisados en la literatura. Se debe notar que el registro de calificaciones de actividades académicas es común en todos los cursos, por lo que son datos que se pueden recopilar fácilmente, lo que permite que esta metodología pueda ser aplicada a una gran cantidad de cursos de distintas áreas. De igual manera, la metodología es flexible en cuanto a la selección de la etapa temporal en la que se realizan las predicciones con base en la exactitud que se considere aceptable.

## Futuras líneas de investigación

A pesar de los avances logrados en esta investigación es conveniente mencionar las posibles exploraciones que se pueden realizar en esta área. Primeramente, puede ser de interés realizar estudios empleando técnicas más complejas que las presentadas en esta investigación, tales como la técnica de voto mayoritario, que involucra una votación entre técnicas de aprendizaje automático, como las mostradas en este trabajo, y tomar una decisión con base en lo que decida la mayoría. También, se han utilizado atributos referentes a calificaciones de actividades académicas, no obstante, se pueden emplear, adicionalmente, atributos demográficos o socioeconómicos que puedan influir en el rendimiento académico.

## Referencias

- Alcaraz, R., Martínez, A., Zangróniz, R. y Rieta, J. J. (2020). Predicción temprana del fracaso en una asignatura de electrónica con técnicas de aprendizaje automático. Ponencia presentada en el XIV Congreso de Tecnologías Aplicadas a la Enseñanza de la Electrónica. Porto, del 8 al 10 de julio de 2020. <https://dialnet.unirioja.es/servlet/articulo?codigo=7980475>.
- Altujjar, Y., Altamimi, W., Al-Turaiki, I. y Al-Razgan, M. (2016). Predicting Critical Courses Affecting Students Performance: A Case Study. *Procedia Computer Science* 82, 65-71. <https://doi.org/10.1016/j.procs.2016.04.010>.
- Antelm, A. M., Cacheiro, M. L. y Gil, A. J. (2015). Análisis del fracaso escolar desde la perspectiva del alumnado y su relación con el estilo de aprendizaje. *Educación y Educadores*, 18(3), 471-489. <https://www.redalyc.org/articulo.oa?id=83443150006>.
- Asif, R., Merceron, A., Ali, S. A. y Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113, 177-194. <https://doi.org/10.1016/j.compedu.2017.05.007>.
- Ayala, E., López, R. E. y Menéndez, V. H. (2021). Modelos predictivos de riesgo académico en carreras de computación con minería de datos educativos. *Revista de Educación a Distancia (RED)*, 21(66). <https://doi.org/10.6018/red.463561>.
- Bakhshinategh, B., Zaiane, O. R., ElAtia, S. y Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*, 23(1), 537-553. <https://doi.org/10.1007/s10639-017-9616-z>.

- Contreras, L. E., Fuentes, H. J. y Rodríguez, J. I. (2020). Predicción del rendimiento académico como indicador de éxito/fracaso de los estudiantes de ingeniería, mediante aprendizaje automático. *Formación Universitaria*, 13(5), 233-246. <https://doi.org/10.4067/S0718-50062020000500233>.
- Del Campo, J., Ramos, G., Morales, R. y Baena, M. (2017). Minería de datos educativos para la predicción personalizada del rendimiento académico. Conferencia Internacional de Procesamiento de la Información. Villa Clara, del 23 al 27 de octubre de 2017. <https://riuma.uma.es/xmlui/handle/10630/15477>.
- Durairaj, M. y Vijitha, C. (2014). Educational Data mining for Prediction of Student Performance Using Clustering Algorithms. *International Journal of Computer Science and Information Technologies*, 5(4), 5987-5991. <http://www.ijcsit.com/docs/Volume%205/vol5issue04/ijcsit20140504249.pdf>.
- Garbanzo, G. M. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. *Educación*, 31(1), 43-63. <https://www.redalyc.org/pdf/440/44031103.pdf>.
- Hernández, J., Ramírez, M. y Ferri, C. (2004). *Introducción a la minería de datos*. Madrid, España: Pearson.
- Hussain, S., Dahan, N. A., Ba-Alwi, F. M. y Ribata, N. (2018). Educational Data Mining and Analysis of Students' Academic Performance Using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, 9(2), 447-459. <https://doi.org/10.11591/ijeecs.v9.i2.pp447-459>.
- Juárez, A., Cortés, J. y Coronilla, U. (2014). Aplicación de la inteligencia artificial en la sistematización de procesos educativos. Caso: Sistema de detección de riesgo escolar en ESCOM. *Revista Iberoamericana de Producción Académica y Gestión Educativa*, 1(1), 140-163. <https://www.pag.org.mx/index.php/PAG/article/view/92/140>.
- Kalaivani, S., Priyadarshini, B. y Selva, B. (2017). Analyzing Student's Academic Performance Based on Data Mining Approach. *International Journal of Innovative Research in Computer Science & Technology*, 5(1), 194-197. <https://doi.org/DOI:10.21276/ijrcst.2017.5.1.4>.
- La Red, D., Karanik, M., Giovannini, M. y Pinto, N. (2015). Perfiles de rendimiento académico: un modelo basado en minería de datos. *Campus Virtuales*, 4(1), 12-30.

- <https://redined.educacion.gob.es/xmlui/bitstream/handle/11162/120661/1.pdf?sequence=1&isAllowed=y>.
- Márquez, C., Cano, A., Romero, C. y Ventura, S. (2012). Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. *Applied Intelligence*, 38(3), 315-330. <https://doi.org/10.1007/s10489-012-0374-8>.
- Mueen, A., Zafar, B. y Manzoor, U. (2016). Modeling and Predicting Students' Academic Performance Using Data Mining Techniques. *International Journal of Modern Education and Computer Science*, 8(11), 36-42. <https://doi.org/10.5815/ijmecs.2016.11.05>.
- Osmanbegović, E. y Suljić, M. (2012). Data Mining Approach for Predicting Student Performance. *Journal of Economics and Business*, 10(1), 3-12. <https://www.econstor.eu/handle/10419/193806>.
- Rivera, K. (2021). Modelo predictivo para la detección temprana de estudiantes con alto riesgo de deserción académica. *Revista Innovación y Software*, 2(2), 6-13. <https://revistas.ulasalle.edu.pe/innosoft/article/view/40/37>.
- Sharma, G. y Vishwakarma, S. K. (2017). Analysis and Prediction of Student's Academic Performance in University Courses. *International Journal of Computer Applications*, 160(4), 40-44. <https://doi.org/10.5120/IJCA2017913045>.
- Silva, M. (2011). El primer año universitario. Un tramo crítico para el éxito académico. *Perfiles Educativos*, 33(especial), 102-114. <http://www.scielo.org.mx/pdf/peredu/v33nspe/v33nspea10.pdf>.
- Usman, U. I., Salisu, A., Barroon, A. I. E. y Yusuf, A. (2019). A Comparative Study of Base Classifiers in Predicting Students' Performance based on Interaction with LMS Platform. *FUDMA Journal of Sciences*, 3(1), 231-239.
- Valero, S., Vargas, A. y García, M. (2010). Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los  $k$  vecinos más cercanos. En Prieto, M. E., Dodero, J. M. y Villegas, D. O. (eds.), *Recursos digitales para la educación y la cultura volumen Kaambal*, (pp. 33-39). Mérida, México: Universidad Tecnológica Metropolitana. [http://fcaenlinea.unam.mx/anexos/1566/1566\\_u6\\_act1b.pdf](http://fcaenlinea.unam.mx/anexos/1566/1566_u6_act1b.pdf).

- Witten, I., Frank, E. y Hall, M. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Massachusetts, United States: Morgan Kaufmann Publishers.
- Xing, W., Guo, R., Petakovic, E. y Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, 47, 168-181. <https://doi.org/10.1016/j.chb.2014.09.034>.