



Metodología para predecir la Deserción Estudiantil en el Ciclo Básico de Programas de Pregrado en la Universidad del Valle (Cali, Colombia)

Brayan David Castro Rodríguez
Universidad del Valle, Colombia
Brayan.castro@correounivalle.edu.co

Oscar Alonso Castaño Macana U
Universidad del Valle, Colombia
oscar.castano@correounivalle.edu.co

María Helena Pinzón Cárdenas
Universidad del Valle, Colombia
maria.pinzon@correounivalle.edu.co

Sandra Milena Barona Ramírez
Universidad del Valle, Colombia
barona.sandra@correounivalle.edu.co

Resumen

Este trabajo desarrolla modelos de machine learning para identificar y focalizar poblaciones de estudiantes con riesgo a desertar y establecer estrategias de prevención en la Universidad del Valle (Cali, Colombia). Utilizando datos históricos de ingreso y los tres primeros semestres, se construyeron cuatro modelos de predicción. El modelo inicial alcanzó una exactitud del 64% y una sensibilidad del 62%, mejorando significativamente en los modelos de seguimiento: 74% de exactitud y 70% de sensibilidad después del primer semestre, 79% y 75% después del segundo, y 81% y 78% después del tercero. Los resultados se actualizan periódicamente en un *dashboard* accesible para la comunidad universitaria disponible

con datos desde el 2020 hasta el 2023. Basándose en estas predicciones, se implementó el piloto de “Consejería Estudiantil” dirigido a estudiantes identificados como posibles desertores para brindar un acompañamiento integral. Este enfoque permitió una intervención focalizada y temprana, que aunque enfrenta desafíos como las necesidades tanto de mantener la precisión y actualización constante de los datos, así como promover la corresponsabilidad de los estudiantes ante el acompañamiento permite establecer acciones de forma preventiva. Estas estrategias han demostrado aportar al proceso de formación, destacando la importancia de acompañar a los estudiantes con mayor riesgo de deserción.

Contextualización

Desde la perspectiva individual, la deserción universitaria es un fenómeno complejo y multidimensional en donde los estudiantes no logran alcanzar los objetivos de la educación superior, lo que representa un fracaso personal e institucional (Castaño et al., 2008). La complejidad inherente a este fenómeno sugiere que ninguna definición única puede capturarlo en su totalidad, subrayando la importancia de abordarlo desde diversas perspectivas para diseñar estrategias de prevención efectivas. A nivel institucional, la deserción se cuantifica por el número de estudiantes que abandonan la universidad antes de completar sus estudios y obtener un título. En el contexto estatal, se entiende como el abandono del sistema educativo en su totalidad (Castaño et al., 2004; Tinto, 1989).

El Ministerio de Educación Nacional (MEN), con el aporte de diferentes Instituciones de Educación Superior, ha definido lineamientos desde los planes sectoriales para atender de manera integral y articulada la multicausalidad y complejidad de la deserción, con un enfoque preventivo. Es así como, el Modelo de Gestión de Permanencia y Graduación Estudiantil, propuesto por el MEN, resalta la importancia de la gestión de la información para mejorar los procesos de acompañamiento y orienta sobre el diseño de propuestas de intervención basadas en modelos contextualizados.

En la búsqueda del éxito académico, donde confluyen dimensiones individuales, familiares, universitarias, económicas y académicas de los estudiantes (Pereira y Vidal, 2020), la Universidad del Valle ha implementado diversos estudios y estrategias para afrontar la problemática de la deserción. A pesar de estos esfuerzos, las intervenciones actuales generalmente se fundamentan en presunciones sobre la población en riesgo de abandonar sus estudios,

destacando la necesidad de desarrollar herramientas más eficientes para identificar a estos estudiantes. En respuesta a esto, la Universidad del Valle creó la Dirección de Desarrollo Estudiantil y Éxito Académico (DEXIA) para articular todas las acciones dirigidas a reducir la deserción institucional y el rezago académico y en conjunto con el Observatorio de Éxito Académico, Permanencia y Graduación se han enfocado en generar herramientas para la toma de decisiones relacionadas con la deserción y orientar a las estrategias de acompañamiento.

Dentro del presente trabajo, se presenta una metodología avanzada para la construcción de modelos clasificadores de deserción en el ciclo básico utilizando técnicas de *machine learning*. Esta metodología se estructura en dos etapas fundamentales. La primera etapa consiste en un modelo inicial que analiza las características de los estudiantes al momento de su ingreso a la Universidad y la segunda etapa involucra modelos de seguimiento que combinan estas características iniciales con los resultados académicos en cada semestre; lo que permite una observación continua desde el ingreso de los estudiantes y a lo largo de los primeros tres semestres, actualizando las predicciones de deserción semestralmente.



Objetivo general

Desarrollar y validar modelos de machine learning para identificar y predecir con precisión los casos potenciales de deserción estudiantil en la Universidad del Valle, utilizando datos históricos y variables relevantes al momento del ingreso y durante los primeros tres semestres.

Objetivos específicos

1. Desarrollar herramientas avanzadas para la detección temprana de riesgos de deserción estudiantil y facilitar la

implementación de estrategias efectivas de retención.

2. Garantizar la disponibilidad continua y actualización periódica de los modelos desarrollados, permitiendo análisis detallados a nivel de sede, facultad o programa.
3. Asegurar el acceso a datos estructurados y oportunos para que las autoridades universitarias puedan implementar acciones correctivas basadas en información sustentada.

Descripción de la Experiencia

Metodología

Datos

Para este estudio se utilizaron datos estructurados en tres bloques de información, provenientes de la División de Registro Académico de la Universidad del Valle y de DEXIA. Estos bloques ofrecen una visión integral de las condiciones y el desempeño de los estudiantes a lo largo de su ciclo académico. Los datos abarcan el período desde el semestre 2015-1 hasta el semestre 2024-1. El primer bloque incluye variables que caracterizan a los estudiantes al momento de su ingreso a la universidad, tales como identificación, datos personales, características demográficas (sexo, edad, lugar de procedencia), programa académico, y resultados de la prueba Saber 11 realizada por el ICFES (Instituto Colombiano para la Evaluación de la Educación).

El segundo bloque contiene información detallada sobre el historial académico de los estudiantes en cada semestre, incluyendo registros de asignaturas matriculadas, calificaciones,

promedios semestrales, estado de las asignaturas (aprobada, reprobada, cancelada o repetida), número de créditos matriculados y aprobados, e indicadores de Bajo Rendimiento Académico (BRA). El tercer bloque, proporcionado por DEXIA, registra el estado de deserción, identificando a los estudiantes que desertaron en el ciclo básico y aquellos que no lo hicieron.

Para el modelo de predicción al momento de la admisión, se consideran variables como edad, sexo, condición de foráneo, puntajes de la prueba Saber 11, y un índice institucional basado en Análisis de Componentes Principales (ACP) que indica el nivel de riesgo de deserción en los programas académicos. El modelo de seguimiento, además de las variables del modelo al momento de la admisión, incluye el promedio académico de los semestres cursados, incidencia de BRA, número de créditos matriculados y aprobados, cantidad de asignaturas canceladas, y porcentaje de créditos repetidos.

Modelos Predictivos

La metodología utilizada en el estudio abarca varias etapas fundamentales. Primero, se desarrolla un modelo inicial que analiza las características de los estudiantes al momento de su ingreso a la universidad. En la segunda etapa,

se implementan modelos de seguimiento que integran estas características iniciales con los resultados académicos obtenidos semestralmente. Este proceso culmina en la creación de modelos predictivos aplicados para predecir la deserción estudiantil en el ciclo básico universitario, como se muestra en la Gráfica 1.

Gráfica 1. Proceso para construcción de los modelos.



En la creación de modelos de machine learning, se propone el uso de técnicas de balanceo de datos y la evaluación de múltiples clasificadores, priorizando la sensibilidad para identificar verdaderos deserción. La selección de características se realiza mediante el criterio de información mutua y el uso de LazyPredict para una rápida selección de algoritmos (Gu et al., 2021; Yang, 2020, Mohamad et al., 2022, Jonas et al., 2021). Se emplean técnicas como SMOTE y Random UnderSampler para abordar el desbalance de clases, siguiendo a Neema y Mduma (2023). Además,

se destacan métricas como precisión, sensibilidad, especificidad y F1-score en la evaluación de los modelos, garantizando así una robusta predicción de la deserción estudiantil.

Para garantizar la efectividad de los modelos en la predicción de la deserción estudiantil, se identificó los riesgos de sobreajuste y subajuste. Posteriormente, se realizó la validación cruzada y a la selección de hiperparámetros mediante técnicas de búsqueda como grid search o randomized search para mitigar estos riesgos (Max, W., y Werner, B., 2020; Umi, Nur,

Hayati., Yudhistira, Arie, Wijaya., 2023). La interpretación del modelo es fundamental para comprender cómo las características impactan en la predicción de la deserción, y se destaca la importancia de proyectar la posible deserción de estudiantes en el ciclo básico utilizando el modelo seleccionado, actualizándose y pronosticando continuamente para nuevas cohortes estudiantiles. Además, se desarrolla un script en Python que permite cargar datos, aplicar modelos predictivos y generar etiquetas de deserción. Este script integra el modelo estimado y realiza un preprocesamiento de datos antes de aplicarlo a nuevos estudiantes.

Aplicación metodológica

Al aplicar la metodología para la construcción de los modelos de *machine learning*, se obtienen cuatro modelos de predicción de deserción en el ciclo básico. El primer modelo, con una exactitud del 64% y una sensibilidad del 62%, identifica la población objetivo al momento de ingreso a la universidad. Este modelo resalta un aporte significativo en la predicción principalmente del índice institucional del programa académico al que pertenece el estudiante y el índice académico relacionado con los resultados de las pruebas Saber 11.

Los otros tres modelos están diseñados para el seguimiento de los estudiantes y se basan en la incorporación de variables adicionales en cada semestre cursado. Estas variables incluyen el promedio académico del estudiante, los porcentajes de créditos aprobados en el semestre anterior y la identificación de bajos rendimientos académicos en el mismo período. Estos modelos muestran una mejora significativa en las métricas de exactitud y sensibilidad en comparación

con el modelo de ingreso: 74% de exactitud y 70% de sensibilidad (modelo de seguimiento primer semestre), 79% de exactitud y 75% de sensibilidad (modelo de seguimiento segundo semestre) y 81% de exactitud y 78% de sensibilidad (modelo de seguimiento tercer semestre). Esta mejora subraya la importancia de enriquecer los modelos con datos académicos actualizados, lo que permite una mejor identificación y seguimiento de los estudiantes en riesgo de deserción.

Los resultados de estos modelos se actualizan con información de cada periodo académico se presentan en un *dashboard*⁴ disponible para la consulta de la comunidad universitaria para los años desde el 2020 hasta el 2023 (ver Gráfica 2), donde se pueden realizar análisis a nivel de sede, facultad o programa sobre las cantidades de posibles deserción.

A partir de los resultados y predicciones obtenidos de los modelos, la DEXIA implementó un piloto de estrategia de retención estudiantil a través del programa de “Consejería Estudiantil”. Este programa se centra en proporcionar un acompañamiento integral a la población objetivo identificada por los modelos predictivos, brindando apoyo personalizado desde el inicio de su trayectoria universitaria. La Consejería Estudiantil abarca diversas dimensiones del desarrollo del estudiante, incluyendo aspectos personales, sociales y académicos, con el objetivo de atender de manera temprana y efectiva las necesidades de los estudiantes en riesgo de deserción.

4 Para observar el dashboard se pueden dirigir al siguiente link: <https://dexia.univalle.edu.co/fuentes-de-informacion>

Gráfica 2. Dashboard de posibles desertores en la Universidad del Valle con el modelo de admitidos.



El acompañamiento temprano, coordinado con las direcciones de programa, facilita la detección y mitigación de problemas potenciales que puedan afectar el rendimiento y la continuidad académica de los estudiantes. Además, este enfoque proactivo permite la implementación de políticas y estrategias específicas para

reducir las tasas de deserción en la institución, al trabajar estrechamente con los estudiantes, la consejería estudiantil no solo promueve su éxito académico, sino que también contribuye a su bienestar integral, creando un entorno favorable para su desarrollo y permanencia.

Resultados

Con lo anterior, se llevó a cabo un ejercicio piloto durante un periodo académico, en el cual se atendió a una población de 133 estudiantes a través de las estrategias de acompañamiento. De estos, 37 fueron identificados por el modelo de ingreso y 96 por los modelos de seguimiento. Aproximadamente el 4% de los estudiantes presentaron dificultades y cayeron en condición de bajo rendimiento académico, mientras que

cerca del 37% no incurrieron en situaciones de dificultades académicas y avanzaron adecuadamente en su proceso de formación.

Estos resultados destacan, en primer lugar, la capacidad de los modelos de *machine learning* para identificar de manera precisa a los estudiantes en riesgo de deserción y permiten una intervención temprana y focalizada. En

segundo lugar, la implementación del programa de Consejería Estudiantil demostró ser efectiva en apoyar a los estudiantes no solo académicamente, sino también en aspectos personales y sociales, lo cual es crucial para su éxito integral. Sin embargo, también se identificaron desafíos relevantes, uno de los principales retos

es la necesidad de mantener actualizados y precisos los datos utilizados por los modelos, lo cual requiere una coordinación constante entre diferentes instancias de la universidad. Además, de desarrollar estrategias adicionales para aquellos estudiantes que, a pesar del apoyo, continúan enfrentando dificultades.

Conclusiones

Los modelos de *machine learning* desarrollados para predecir la deserción en el ciclo básico universitario demostraron tener una alta precisión. El modelo de ingreso alcanzó una exactitud del 64% y una sensibilidad del 62%, mientras que los modelos de seguimiento, que incorporaron variables adicionales de desempeño académico, lograron una exactitud de hasta el 81% y una sensibilidad del 78%. Esta exactitud destaca la capacidad de estas herramientas para identificar de manera precisa a los estudiantes en riesgo de deserción.

La importancia del seguimiento continuo es evidente en los resultados de los modelos de seguimiento, que mejoraron significativamente las métricas de predicción en comparación con el modelo de ingreso. Al incorporar datos académicos actualizados semestralmente, estos modelos permiten una identificación más precisa de los estudiantes en riesgo. Esto subraya la necesidad de enriquecer continuamente los modelos con información actualizada para mantener y mejorar la exactitud del pronóstico.

El programa de “Consejería Estudiantil” con su intervención temprana y focalizada, basada en las predicciones de los modelos, proporcionó un apoyo integral que favoreció su desarrollo académico, personal y social, con

aproximadamente el 37% de los estudiantes no incurrieron en situaciones de dificultades académicas y avanzando adecuadamente en su formación.

El mantener actualizados los datos utilizados por los modelos predictivos se ha convertido en un desafío continuo que requiere una estrecha coordinación entre diferentes instancias de la universidad. La necesidad de actualización constante es fundamental para garantizar la efectividad de los modelos y la precisión de las predicciones, asegurando que los estudiantes en riesgo sean identificados y apoyados de manera oportuna.

A pesar de la efectividad del programa de consejería, aproximadamente el 4% de los estudiantes continuaron enfrentando dificultades académicas, como incurrir nuevamente en bajo rendimiento. Esto resalta la necesidad de desarrollar estrategias adicionales y más específicas para aquellos estudiantes que, a pesar del apoyo temprano, siguen teniendo problemas en su trayectoria académica. La identificación y el apoyo continuos son esenciales para abordar de manera integral el problema de la deserción universitaria.



Referencias

- Castaño, E., Gallón, S., Gómez, K., y Vásquez, J. (2004). Deserción estudiantil universitaria: una aplicación de modelos de duración. *Lecturas de Economía*, 60, pp. 41-65.
- Castaño, E., Gallón, S., Gómez, K., y Vásquez, J. (2008). Análisis De Los Factores Asociados A La Deserción Estudiantil En La Educación Superior: Un Estudio De Caso. *Revista de Educación*, 345, pp. 255-280.
- Jonas, M. Hanselle, A. Tornede, M., Dominik, W y Eyke, H. (2021). Algorithm Selection as Superset Learning: Constructing Algorithm Selectors from Imprecise Performance Data. doi: 10.1007/978-3-030-75762-5_13
- Max, W y Werner, B. (2020). Evaluation of multiple prediction models: A novel view on model selection and performance assessment.. *Statistical Methods in Medical Research*, doi: 10.1177/0962280219854487
- Mduma, N. (2023). Data Balancing Techniques for Predicting Student Dropout Using Machine Learning. *Data*, doi: 10.3390/data8030049
- Mohamad, A., Sim, K. y Hart, E. (2022). Automated Algorithm Selection: from Feature-Based to Feature-Free Approaches. *Journal of Heuristics*, doi: 10.1007/s10732-022-09505-4
- Nur, U., Yudhistira, H. y Wijaya, A. (2023). Analisa terjadinya overfitting dan underfitting pada algoritma naive bayes dan decision tree dengan teknik cross validation. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1):262-269. doi: 10.36040/jati.v7i1.6329
- Pereira, A. y Vidal, M. (2020). Deserción estudiantil en la educación superior: reflexiones sobre la gestión enfocada en la retención o la permanencia. *Revista Educación*, 45(1), pp. 519-533. <https://doi.org/10.15517/revedu.v45i1.40602>
- Tinto, V. (1989). Definir la deserción: una cuestión de perspectiva. *Revista de La Educación Superior*, 71, pp. 1-9.