



Modelo de clasificación de asignaturas y detección de atipicidades en indicadores académicos de la Universidad del Valle

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

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

Diego Fernando Muñoz Muñoz
Universidad del Valle, Colombia
diego.munoz.munoz@correounivalle.edu.co

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

Resumen

En el presente trabajo se utilizaron datos de registros de asignaturas matriculadas por estudiantes entre 2008-1 y 2023-2 de la Universidad del Valle en Colombia, con los cuales se calcularon cuatro indicadores: tasa de cancelación, tasa de repitencia, promedio de calificación y tasa de reprobación que fueron relacionados mediante el Análisis de Componentes Principales (ACP) con el fin de reducir la dimensionalidad y generar un índice compuesto que prioriza las asignaturas según su desempeño. Además, se implementó una metodología para detectar situaciones atípicas mediante el Método del Rango Intercuartílico (IQR) para identificar registros con desviaciones significativas de la norma. Los resultados de estos modelos han facilitado el reconocimiento de patrones y tendencias en las asignaturas que

históricamente han presentado una dificultad particular en los estudiantes, generando problemas de abandono o deserción estudiantil. De esta manera se logró facilitar la orientación para la toma de decisiones académicas y a establecer acciones correctivas que se espera se reflejen en los procesos de aprendizaje del estudiante. A partir de esos resultados se desarrollaron visualizadores interactivos, que han sido herramientas para que profesores y directivos interpreten datos y se definan oportunidades de mejora en el proceso de enseñanza-aprendizaje.

Palabras clave: Indicadores Académicos, Clasificación de Asignaturas, Análisis de Componentes Principales, Deserción Estudiantil.

Contextualización

La Universidad del Valle, ubicada en Colombia, comparte la misión de las demás instituciones de educación superior de formar mediante la generación, transformación, aplicación y difusión del conocimiento en los ámbitos de las ciencias, la técnica, la tecnología, las artes, las humanidades y la cultura en general. Acorde con lo establecido por Escobar et al. (2021) la expansión de la cobertura educativa ha dificultado la detección temprana de situaciones que impiden el éxito académico, la identificación oportuna de aspectos que influyen en el proceso de enseñanza-aprendizaje es determinante para implementar estrategias de apoyo y mejora continua en los procesos educativos. Es así como, desde la Dirección de Desarrollo Estudiantil y Éxito Académico (DEXIA) se ha propuesto desarrollar herramientas para el análisis que permitan a docentes y directivos orientar los procesos para mitigar las dificultades que presentan los estudiantes en las respectivas asignaturas e implementar estrategias, como fortalecer el programa de tutorías/monitorias y estudiar las metodologías de enseñanza-aprendizaje, entre otras.

La motivación para el desarrollo de las herramientas estableció dos escenarios, uno de ellos relacionado con la importancia de implementar un modelo capaz de determinar problemas estructurales en las asignaturas por programa académico, es decir, realizar una caracterización histórica de los principales indicadores académicos por asignatura (tasa de

cancelación, tasa de repitencia, promedio de calificación y tasa de reprobación), para así identificar los patrones de comportamiento y priorizar las asignaturas más críticas. El segundo escenario, considera la necesidad de contar con un modelo que permita realizar un monitoreo o seguimiento semestral e histórico a cada una de las asignaturas, facilitando la identificación de comportamientos atípicos en los indicadores académicos. De esta manera, al identificar las asignaturas que han presentado mayor dificultad a los estudiantes, es posible generar estrategias que faciliten el aprendizaje de estos en dichas asignaturas, estimulando así que puedan culminar de forma exitosa su vida universitaria, disminuyendo así los escenarios de abandono estudiantil.

Como resultado, se diseñaron modelos que se fundamentan en el ACP y técnicas univariantes de detección de outliers, que han demostrado ser efectivas en estudios previos para evaluar el desempeño académico y detectar anomalías (Escobar et al., 2021; Arifin, M. et al, 2023; Almardeny, Y. et al, 2022; Castaño, O. et al, 2019). La implementación de estos modelos no solo caracteriza y priorizan las asignaturas que presentan mayores desafíos, sino también monitorean de manera continua el comportamiento de los indicadores académicos para facilitar la identificación de situaciones atípicas y orientar la toma de decisiones oportunas. Esta estrategia integral busca mejorar la interacción entre estudiantes, docentes y directivos.

Objetivos

El trabajo está orientado a proporcionar información mediante dashboards para el análisis continuo de datos académicos y facilitar la visualización a docentes, directivos y partes

interesadas, la clasificación de asignaturas e indicadores académicos asociados. Esto permite, en la toma de decisiones, la identificación de las asignaturas con alta prioridad y la detección

de anomalías en periodos específicos como parte del seguimiento y evaluación del progreso estudiantil a niveles grupales o individuales. De esta manera, se espera que las decisiones o estrategias puedan ser focalizadas en dichas

asignaturas que presentan mayor dificultad por parte de los estudiantes, facilitando la absorción de conocimiento y estimulando la culminación exitosa de su etapa universitaria.

Descripción de la Experiencia

Datos

Los datos utilizados en este estudio corresponden a los registros de las asignaturas matriculadas por alrededor de 30 mil estudiantes de la Universidad del Valle semestralmente, desde el periodo 2008-1 hasta el 2023-2. Estos registros están organizados por asignatura y contienen información detallada que permite identificar en qué sede, facultad y programa se realizó la matrícula. Cada registro incluye variables como la calificación obtenida por el estudiante en la asignatura, si el estudiante reprobó la asignatura, si canceló la asignatura y si está viendo la asignatura por primera vez o si la está repitiendo.

A partir de estos datos, se asociaron las cuatro variables de interés para cada asignatura: las tasas de cancelación, repitencia y reprobación que representan el porcentaje de estudiantes que cancelaron, repitieron y reprobaron la asignatura, respectivamente. Así como, el promedio de calificación que refleja la nota media obtenida por los estudiantes en la asignatura.

Aplicación Metodológica

Caracterización de las Asignaturas

Para diseñar la herramienta de caracterización de las asignaturas, se llevó a cabo un estudio detallado de cuatro indicadores académicos: las tasas promedio de reprobación, cancelación y

repitencia, así como el promedio de calificación de las asignaturas (Caballero, L. et al, 2011). Estos indicadores se seleccionaron debido a su relevancia en la evaluación del desempeño académico y su capacidad para reflejar las dificultades que enfrentan los estudiantes en diversas asignaturas.

El ACP se utilizó como metodología para analizar el comportamiento de los indicadores a lo largo del tiempo y reducir la dimensionalidad de las variables, permitiendo así una interpretación clara y manejable de los datos. La reducción de dimensionalidad se llevó a cabo mediante la estimación de coeficientes que capturan la correlación existente entre los cuatro indicadores.

El resultado del ACP fue la generación de un índice compuesto que prioriza las asignaturas en función de los valores obtenidos. Este índice se construyó de manera que aumenta a medida que aumentan las tasas de reprobación, cancelación y repitencia, y disminuye ante aumentos en las calificaciones promedio de los estudiantes (Becerra, 2010; Castrillón-Gómez et al, 2020; Pardo, 2020). De esta forma, el índice proporciona una medida integrada que refleja de manera efectiva el rendimiento académico de las asignaturas.

Posteriormente, se procedió a clasificar las asignaturas en tres categorías dependiendo de su prioridad como: alta, media y baja, basándose en el índice obtenido. Esta

clasificación se determinó según cuántas desviaciones estándar se alejan de la media del índice: una asignatura se clasifica como “alto” si su índice está una desviación estándar por encima de la media, como “bajo” si está una desviación estándar por debajo de la media, y como “medio” si se encuentra entre estas desviaciones estándar.

La clasificación se realiza teniendo en cuenta las asignaturas correspondientes a un programa específico en cada sede. Esto significa que una misma asignatura puede tener diferentes clasificaciones dependiendo de la sede en la que se imparte, ya que los comportamientos y rendimientos académicos pueden variar entre sedes.

Esta metodología permitió una evaluación estructurada y objetiva de las asignaturas, proporcionando una herramienta pertinente al momento de identificar aquellas asignaturas que presentan mayores desafíos y orientar así la implementación de estrategias educativas específicas para mejorar el rendimiento académico y la experiencia de aprendizaje de los estudiantes.

Indicadores de Situaciones Atípicas en las Asignaturas

Una metodología basada en técnicas univariantes de detección de *outliers* se eligió para monitorear o supervisar automáticamente las tendencias y su desviación del comportamiento histórico de los cuatro indicadores académicos por asignatura (la calificación promedio y las tasas de reprobación, repitencia y graduación). Los *outliers* son los datos o registros que presentan una desviación significativamente elevada del conjunto de datos.

El Método del Rango Intercuartílico (IQR) se implementó como técnica de detección de *outliers* y se estableció el rango intercuartílico

como la diferencia entre el valor asociado al percentil 75 más alto y el percentil 25 del conjunto de datos seleccionado, se determinaron como *outliers* aquellos registros cuyo valor fue superior al valor asociado al percentil 75 más el IQR multiplicado por una constante o inferior al valor asociado al percentil 25 menos el IQR multiplicado por una constante. El valor de la constante que se multiplica por el IQR refleja la medida de ajuste o grado de tolerancia permitido al momento de realizar la clasificación de atipicidad, de manera que, a valor mayor de la constante, los registros se distancian en mayor medida para ser clasificados como atípicos (Geetha et al, 2023).

El principal desafío en la implementación de metodologías de supervisión y control en la Universidad del Valle radica en la comparabilidad de datos debido a la heterogeneidad entre sus diferentes sedes, facultades, programas académicos y asignaturas. Esta diversidad limita la comparación efectiva de los indicadores académicos. Para abordar esta complejidad, se establece que los indicadores académicos sólo son comparables con sus propios registros históricos, permitiendo una evaluación más precisa adaptada a cada contexto académico, identificando tendencias y patrones específicos dentro de cada grupo y adaptando estrategias de mejora a las particularidades de cada segmento.

Se establece el término “partición” para la elección de la base o conjunto de registros del análisis comparativo de los indicadores académicos de la asignatura y se refiere a la combinación única de las categorías o grupos asociados a cada indicador académico; asignatura, programa académico, facultad y sede. De esta forma se compara el valor de cada uno de los cuatro indicadores académicos de interés con el registro histórico presentado por dicho indicador en la partición perteneciente.

El criterio de marcación se genera a partir del cálculo del valor crítico por partición para cada indicador académico bajo el método de rango intercuartílico, que se calculó con la siguiente ecuación:

$$VC_i = P_{75, i} + (k)IQR_i$$

Donde VC representa el valor crítico calculado para la i -ésima partición, $P_{75, i}$ es el percentil 75, k es el multiplicador del rango intercuartílico que permite modificar el grado de atipicidad permitido IQR_i es el rango intercuartílico de la i -ésima partición. El criterio de marcación consta de clasificar como alerta las asignaturas cuyo valor del indicador académico registrado en un semestre en determinado programa académico, facultad y sede, supere el valor crítico calculado.

Experiencia de la Aplicación Metodológica

El desarrollo de los modelos de caracterización de las asignaturas e indicadores de situaciones atípicas en las asignaturas han aportado a las directivas institucionales, en su misión académica, a promover el seguimiento del desempeño académico y generar análisis orientadores para la toma de decisiones que contribuyan a mitigar la problemática del fracaso académico. Esto se enmarca en un contexto de necesidad creciente de supervisión, orientación y apoyo en el proceso de enseñanza-aprendizaje, con el fin de aportar continuamente a una educación de calidad. La implementación de estos modelos abarca todas las 9 sedes, 10 facultades y 81 programas de la Universidad del Valle, proporcionando un enfoque integral que permite una evaluación detallada y precisa de las condiciones académicas en la institución y la cual permite articularse con otras instancias y herramientas para un mayor alcance. Los usuarios

directos de los resultados de los modelos son los profesores, los directores de programa y la DEXIA quienes interactúan a través de visualizadores diseñados específicamente para facilitar la interpretación de los datos.

Estos visualizadores interactivos, desarrollados para cada modelo de análisis, ofrecen a los usuarios una vista personalizada y detallada de los indicadores académicos relevantes, permitiendo filtrar la información por asignatura, sede, facultad y programa. Esto facilita la identificación rápida de áreas problemáticas y oportunidades de mejora para los profesores y directores de programa, mientras que los estudiantes se benefician indirectamente de las acciones preventivas y correctivas derivadas del análisis de los datos. La ejecución de esta iniciativa se ha llevado a cabo desde el periodo 2023-2 hasta el 2024-1, con un enfoque en la recolección y análisis continuo de datos acumulados desde el periodo 2008-1 hasta el periodo 2023-2, proporcionando una herramienta oportuna para acceder a información académica específica y detallada.

Además, desde DEXIA se han organizado una serie de talleres destinados a capacitar a los profesores y directores de programa en el uso de estas herramientas. Estos talleres están diseñados para asegurar que los usuarios puedan interpretar los datos de manera efectiva y oportuna, identificando así oportunidades de mejora y tomando decisiones informadas para favorecer los procesos de enseñanza-aprendizaje.

Resultados

Los resultados de la implementación de estos modelos han sido significativos en varios aspectos, en primer lugar, la caracterización de las asignaturas ha permitido identificar patrones y tendencias en indicadores clave como las tasas de cancelación, repitencia, reprobación y promedios de calificación; estos datos proporcionan una base sólida para comprender las dinámicas académicas en distintas asignaturas y programas. La implementación de la metodología de ACP genera una ventaja particular que es la posibilidad de generar una priorización a partir del patrón de comportamiento de los indicadores académicos. La última actualización del modelo de caracterización se realizó al incluir los registros del semestre 2023-2, los resultados de este análisis muestran el 27,2% de las asignaturas con una prioridad “alta”, 32% fue clasificada como prioridad “media” y el 40,8% con prioridad “baja”.

El modelo de indicadores de situaciones atípicas ha sido particularmente útil para detectar asignaturas que presentan desviaciones del comportamiento normal, tanto en términos

positivos como negativos. Los visualizadores desarrollados para ambos modelos han facilitado la accesibilidad y comprensión de los datos, permitiendo a los profesores y directores de programa identificar rápidamente las áreas que requieren atención. En la Gráfica 1, se observan los resultados del modelo de detección de indicadores atípicos agrupados para la institución en función de los períodos académicos, desde el periodo 2015-1 hasta el periodo 2023-2, y del número de indicadores académicos por asignatura que han sido identificados como atípicos. Antes de la pandemia ocasionada por el virus COVID-19 se observa que el número de atípicos detectados presenta un comportamiento estable con una leve tendencia positiva, durante la pandemia se reduce drásticamente el número de alertas debido a la anormalidad académica ocasionada por el confinamiento obligatorio y las medidas excepcionales tomadas por las IES. Posterior al periodo de confinamiento el número de alertas presenta una aceleración hasta alcanzar el punto máximo de indicadores atípicos en un semestre (1243 indicadores).

Gráfica 1. Número de indicadores académicos atípicos identificados por periodo académico.

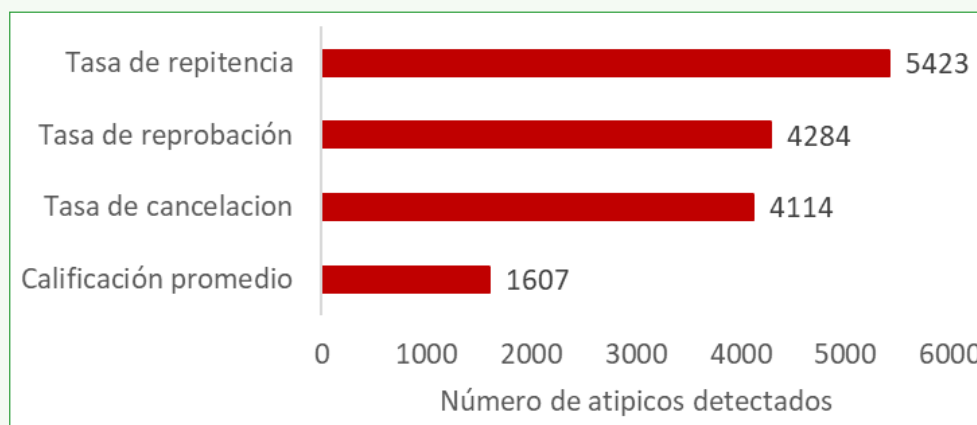


En la gráfica 2, se presenta el número de indicadores atípicos identificados desde el año 2015 a 2023, donde el indicador académico con más alertas generadas es la tasa de repitencia que agrupa el 35,2% de las alertas generadas, seguido por la tasa de reprobación (27,8%), tasa de repitencia (26,7%) y calificación promedio (10,4%). El resultado del promedio por programa académico de indicadores atípicos para cada facultad evidencia que los programas académicos de la Facultad de Salud presentan el mayor número de indicadores atípicos identificados, dado que en promedio sobre cada programa académico de dicha facultad se han detectado 162 alertas, seguida por la Facultad

de Ingeniería con 141 alertas y los programas de la Facultad de Artes integradas con un promedio de 110 alertas.

Entre los logros más relevantes de esta práctica educativa se destacan la mejora en la toma de decisiones académicas, la identificación oportuna de asignaturas problemáticas y la promoción de un enfoque basado en datos para la gestión académica. Los desafíos, por otro lado, han incluido la necesidad de una capacitación continua para los usuarios de las herramientas y la adaptación de las estrategias de enseñanza-aprendizaje y de evaluación en respuesta a los hallazgos de los modelos.

Gráfica 2. Número de situaciones atípicas por tipo de indicador desde 2015 hasta 2023.



Conclusiones

La implementación de modelos de análisis y monitoreo de indicadores académicos en la Universidad del Valle ha permitido la comprensión de las dinámicas académicas y la identificación de aspectos críticos en las asignaturas. Los modelos desarrollados, basados en el ACP y técnicas de detección de *outliers*, han facilitado la clasificación de asignaturas

y la detección de situaciones atípicas, proporcionando información valiosa para la toma de decisiones y la implementación de alternativas en pro del éxito en el proceso de formación de los estudiantes. La caracterización de las asignaturas ha revelado patrones significativos en indicadores clave como calificaciones promedio y tasas tanto de reprobación, como

de cancelación y de repitencia lo que ha permitido una priorización de asignaturas críticas. Asimismo, el monitoreo de indicadores atípicos ha generado alertas que han facilitado la implementación de acciones correctivas, mejorando así la supervisión académica.

Los visualizadores interactivos han sido esenciales para la interpretación de los datos, permitiendo a docentes y directivos identificar rápidamente asignaturas que requieren atención. Las experiencias obtenidas evidencian una identificación oportuna de asignaturas problemáticas y la promoción de un enfoque basado en datos para la gestión académica. A pesar de los desafíos en la capacitación continua de usuarios y la adaptación de estrategias de

enseñanza, la experiencia ha demostrado ser una herramienta eficaz para aportar a la mejora en el desempeño académico y la experiencia de aprendizaje de los estudiantes. No obstante, al momento de presentar esta investigación, la frecuencia de actualización de los resultados del modelo es semestral, lo cual presenta limitaciones respecto a que tan oportunas son las alertas que se puedan generar para la prevención de resultados adversos. El poder contar con información en tiempo real o con mayor frecuencia de actualización de los indicadores podría generar mejores resultados respecto a estrategias o decisiones que se puedan tomar en pro del éxito académico de los estudiantes en la etapa de educación superior.

Referencias

- Almardeny, Y., Boujnah, N., y Cleary, F. (2022). A novel outlier detection method for multivariate data. *IEEE Transactions on Knowledge and Data Engineering*. <https://doi.org/10.1109/tkde.2020.3036524>
- Arifin, M., Widowati, W., Farikhin, F., y Gudnanto, G. (2023). A regression model and a combination of academic and non-academic features to predict student academic performance. *TEM Journal*, 12(2). <https://doi.org/10.18421/tem122-31>
- Becerra, M. (2010). Comparación del análisis factorial múltiple (AFM) y del análisis en componentes principales para datos cualitativos (Prinqual), en la construcción de índices. Tesis de Maestría en Estadística. Universidad Nacional de Colombia.
- Caballero, L. B., Castillo, Y. B., y Álvarez, C. B. (2011, November). Comparación de las tasas de aprobación, reprobación, abandono y costo estudiante de dos cohortes en carreras de Licenciatura en Ingeniería en la Universidad Tecnológica de Panamá. *Actas del I CLABES*.
- Castaño, O., Arce, J., y Ochoa, A. (2019). Caracterización y construcción de índices de riesgo de deserción: Estudio aplicado en la Facultad de Ingeniería de la Universidad del Valle. *Actas del IX CLABES*. Recuperado de: <https://revistas.utp.ac.pa/index.php/clabes/issue/view/122/154>
- Castrillón-Gómez, O., Sarache, W., y Ruiz-Herrera, S.. (2020). Predicción de las principales variables que conllevan al abandono estudiantil por medio de técnicas de minería de datos. *Formación universitaria*, 13(6), pp. 217-228. <https://dx.doi.org/10.4067/S0718-50062020000600217>
- Escobar, J., Castaño, O., y Castro, B. (2021). Patrón estructural de la deserción en la Universidad del Valle. *Actas del X CLABES*. Recuperado a partir de; <https://revistas.utp.ac.pa/index.php/clabes/article/view/3350>
- Geetha, A. y Sangeetha, T. (2023). Outlier Detection in Single Universal Set using Intuitionistic Fuzzy Proximity Relation based on A Rough Entropy-Based Weighted Density Method. *International Research Journal on Advanced Science Hub*, doi: 10.47392/irjash.2023.s067
- Pardo, C. (2020). Estadística descriptiva multivariada. Universidad Nacional de Colombia. Recuperado a partir de: <https://repositorio.unal.edu.co/handle/unal/79914>