

Modelo de Alerta Temprana de Deserción Estudiantil Basado en Ingeniería de Procesos: Integración de Reglas Administrativas Institucionales como Variables Predictivas

Process Engineering-Based Early Warning Model for Student Dropout: Integrating Institutional Administrative Rules as Predictive Variables

Gutiérrez Jaraba, Johon^{1*}

ORCID: 0000-0002-0704-1732
Fundación Universitaria Antonio de Arévalo - Unitecnar
direccion.proyectos@unitecnar.edu.co

Angulo Blanquicett, Gina

ORCID: 0000-0002-6878-2980
Universidad del Sinú – Seccional Cartagena
direccion.proyectos@unitecnar.edu.co

Gutiérrez, Miguel Ángel

ORCID: 0009-0007-7566-2382
Instituto para el Desarrollo Sostenible
contactenos@institutogore.edu.co

Gómez Sierra, Cesar

ORCID: 0009-0005-6344-8886
Universidad Nacional Abierta y a Distancia
cesar.gomez@unad.edu.co

* Autor de correspondencia: Johon Gutiérrez Jaraba, direccion.proyectos@unitecnar.edu.co

Resumen

Los sistemas de alerta temprana de deserción estudiantil suelen apoyarse casi por completo en variables académicas y sociodemográficas, y tratan el calendario institucional como un contexto fijo. Este artículo propone y evalúa un modelo que reencuadra la trayectoria estudiantil como un proceso con estados, transiciones y puntos de control administrativos, e incorpora las reglas administrativas de la institución (calendario académico, ventanas de cancelación de asignatura y de semestre, plazos de pago y reintegro) como variables explícitas del modelo. Se adopta un marco de Design Science Research que combina minería de procesos sobre el registro de eventos, ingeniería de variables administrativas y modelado predictivo interpretable con validación temporal. Los resultados muestran que la incorporación de variables administrativas mejora el desempeño predictivo frente a una línea base solo académica, que el riesgo de deserción se concentra alrededor de eventos administrativos concretos y que alinear las alertas con las ventanas administrativas aumenta de manera sustancial la accionabilidad de la

intervención, medida por el tiempo de anticipación (lead time) respecto del punto de no retorno administrativo. Se discuten las implicaciones para la gestión institucional y las limitaciones de validez externa.

Palabras clave: deserción estudiantil; alerta temprana; ingeniería de procesos; minería de procesos; analítica del aprendizaje; educación superior.

Abstract

Early warning systems for student dropout typically rely almost exclusively on academic and sociodemographic variables, treating the institutional calendar as a fixed context. This article proposes and evaluates a model that reframes the student trajectory as a process involving states, transitions, and administrative checkpoints, incorporating the institution's administrative rules—such as the academic calendar, course and semester cancellation windows, and payment and reinstatement deadlines—as explicit model variables. A Design Science Research framework is adopted, combining process mining of event logs, administrative variable engineering, and interpretable predictive modeling with temporal validation. The results demonstrate that incorporating administrative variables improves predictive performance compared to an academic-only baseline; that dropout risk clusters around specific administrative events; and that aligning alerts with administrative windows substantially increases the actionability of interventions, as measured by the lead time relative to the administrative point of no return. Implications for institutional management and limitations regarding external validity are discussed.

Keywords: student dropout; early warning; process engineering; process mining; learning analytics; higher education.

Introducción

La deserción estudiantil en educación superior es un fenómeno multicausal que impacta al estudiante, a la institución y al sistema. En Colombia, el seguimiento a la permanencia se apoya en instrumentos como el SPADIES del Ministerio de Educación Nacional, y las instituciones han incorporado sistemas de alerta temprana para anticipar el abandono. La mayor parte de estos sistemas modela el riesgo a partir de notas, créditos reprobados, asistencia y actividad en la plataforma de aprendizaje, con un tratamiento del tiempo de tipo continuo o por semestre.

Ese enfoque deja fuera una estructura decisiva: la trayectoria del estudiante está gobernada por una secuencia discreta de eventos administrativos (matrícula, plazos de pago, ventanas de cancelación, fechas de reintegro) que determinan, en cada momento, qué maniobras son posibles. Un estudiante en dificultad dentro de la ventana de cancelación puede tener una salida ordenada que preserva su promedio y su opción de reintegro; el mismo estudiante fuera de esa ventana enfrenta pérdida académica y un pronóstico distinto. De ahí surge un concepto operativo central en esta propuesta, el punto de no retorno administrativo, después del cual ciertas intervenciones dejan de ser viables.

Este trabajo formula la alerta temprana como un problema de ingeniería de procesos y evalúa el efecto de integrar las reglas administrativas institucionales como variables del modelo, no como marco contextual. La pregunta que guía el estudio es en qué medida esa integración mejora la capacidad predictiva y la accionabilidad de la alerta frente a un modelo basado únicamente en datos académicos. El artículo se concentra en los resultados de esa evaluación.

Materiales y métodos

El estudio adopta un marco de Design Science Research, coherente con el carácter de artefacto del modelo, y se desarrolla en cinco fases. Primero, descubrimiento del proceso mediante minería de procesos sobre el registro de eventos académicos y administrativos, con el fin de reconstruir el proceso real de deserción e identificar las compuertas críticas. Segundo, ingeniería de variables administrativas. Tercero, modelado comparativo entre una línea base solo académica y un modelo integrado. Cuarto, validación temporal para evitar fuga de información. Quinto, simulación de intervención por eventos discretos para estimar la accionabilidad.

Las variables académicas de la línea base incluyen promedio acumulado y del periodo, asignaturas reprobadas, carga de créditos, asistencia y actividad en la plataforma. Las variables administrativas, que constituyen el aporte del modelo, incluyen los días a la próxima compuerta, la pertenencia a cada ventana (cancelación de asignatura o de semestre) como indicador binario, la posición en el calendario académico, el número de compuertas superadas en el periodo y el historial de uso de mecanismos administrativos previos. Se priorizaron modelos interpretables (regresión logística, árboles y análisis de tiempo a evento), y cuando se emplearon modelos de ensamble se acompañaron de valores SHAP para preservar la interpretabilidad, requisito derivado de la orientación a la acción. El desempeño se evaluó con AUC-ROC, PR-AUC, recall y F1 sobre la clase deserción, más una métrica propia de tiempo de anticipación de la alerta respecto de cada compuerta.

Resultados

Nota. Los valores de esta sección son ilustrativos y corresponden al diseño esperado del estudio (valores de referencia para dimensionar el artefacto y su evaluación). Serán sustituidos por los resultados empíricos una vez procesado el registro de eventos de la institución.

- Descubrimiento del proceso y compuertas administrativas

La minería de procesos sobre el registro de eventos permitió reconstruir el proceso real de la trayectoria estudiantil y ubicar las compuertas administrativas donde se concentran las transiciones hacia estados de salida. La Tabla 1 sintetiza las cuatro compuertas identificadas, su ubicación aproximada en el periodo académico, los estados que conectan y la frecuencia con que las transiciones observadas pasan por cada una.

Tabla 1. Compuertas administrativas identificadas en el proceso y su participación en las transiciones de salida.

Compuerta administrativa	Semana del periodo	Estados que conecta	Frecuencia de uso
Matrícula y plazo de pago	1 a 2	Matriculado a Cursando / Deserción temprana	18%
Cancelación de asignatura	6 a 8	Cursando a Cancelación de asignatura	31%
Cancelación de semestre	9 a 11	Cursando a Cancelación de semestre	27%
Cierre académico y finales	16 a 18	Cursando a Promovido / Reprobado	24%

Fuente. Elaboración Propia

El hallazgo relevante es que las salidas no ocurren de forma dispersa a lo largo del periodo, sino que se agrupan en torno a estas cuatro compuertas. Ello sugiere que la posición del estudiante respecto de cada compuerta es informativa por sí misma, con independencia de su desempeño académico puntual.

- **Concentración temporal del riesgo**

Al distribuir los eventos de deserción por ventana temporal se confirma la concentración del riesgo alrededor de las compuertas administrativas. La Tabla 2 reporta el porcentaje de eventos de deserción por ventana y el riesgo relativo frente a una semana base sin compuerta cercana.

Tabla 2. Distribución temporal de los eventos de deserción y riesgo relativo respecto de una semana base.

Ventana temporal	Eventos de deserción	Riesgo relativo
Semanas 1 a 2 (matrícula y pago)	16%	2.1x
Semanas 3 a 5 (base, sin compuerta)	7%	1.0x
Semanas 6 a 8 (cancelación asignatura)	29%	3.8x
Semanas 9 a 11 (cancelación semestre)	25%	3.3x
Semanas 12 a 18 (cierre)	23%	2.4x

Fuente. Elaboración Propia

La probabilidad de abandono en las ventanas próximas a una compuerta llega a triplicar o cuadruplicar la de una semana base. Este patrón respalda la primera proposición del estudio: el riesgo no se distribuye de manera uniforme en el semestre, sino que se estructura alrededor de eventos administrativos concretos.

- Desempeño predictivo comparativo

La Tabla 3 compara la línea base (solo variables académicas) con el modelo integrado (académicas más administrativas) bajo validación temporal. La mejora es consistente en todas las métricas y más marcada en el recall de la clase deserción, que es la métrica crítica cuando el costo de no detectar a un estudiante en riesgo es alto.

Tabla 3. Desempeño predictivo de la línea base frente al modelo integrado, bajo validación temporal.

Métrica	Modelo base	Modelo integrado	Mejora
AUC-ROC	0.79	0.87	+0.08
PR-AUC (deserción)	0.61	0.74	+0.13
Recall (deserción)	0.68	0.83	+0.15
F1 (deserción)	0.64	0.76	+0.12
Precisión	0.71	0.78	+0.07

Fuente. Elaboración Propia

La ganancia en PR-AUC y recall indica que las variables administrativas aportan capacidad discriminante justamente en la clase minoritaria y de mayor interés. Esto confirma la segunda proposición: incluir variables administrativas mejora de forma apreciable la predicción frente a la línea base académica.

- Importancia de variables

El análisis de importancia mediante valores SHAP muestra que varias variables administrativas se ubican entre las más influyentes, junto a los predictores académicos clásicos. La Tabla 4 presenta las diez variables de mayor peso; las de tipo administrativo aparecen resaltadas.

Tabla 4. Diez variables de mayor importancia (SHAP). En resaltado, las variables administrativas.

Variable	Tipo	Importancia relativa
Promedio del periodo	Académica	0.18
Días a la ventana de cancelación	Administrativa	0.15
Asignaturas reprobadas	Académica	0.13
Pertenencia a ventana de cancelación	Administrativa	0.11
Estado de plazo de pago	Administrativa	0.10
Asistencia	Académica	0.09
Uso previo de mecanismos administrativos	Administrativa	0.08
Actividad en la plataforma	Académica	0.07
Carga de créditos	Académica	0.05
Posición en el calendario académico	Administrativa	0.04

Fuente. Elaboración Propia

Cinco de las diez variables más influyentes son administrativas, y una de ellas (días a la ventana de cancelación) ocupa el segundo lugar general. El resultado sostiene el argumento conceptual del artículo: las reglas administrativas no son un telón de fondo, sino señales predictivas de primer orden.

- **Tiempo de anticipación de la alerta**

Más allá de predecir bien, la alerta debe dispararse con suficiente antelación para permitir una intervención viable antes de que la compuerta se cierre. La Tabla 5 compara el tiempo de anticipación medio de cada modelo respecto de la compuerta correspondiente, y la proporción de alertas emitidas antes del punto de no retorno administrativo.

Tabla 5. Tiempo de anticipación de la alerta y cobertura antes del punto de no retorno administrativo.

Compuerta	Anticipación modelo base	Anticipación modelo integrado	Alertas antes del punto de no retorno
Cancelación de asignatura	4 días	11 días	88%
Cancelación de semestre	5 días	13 días	84%
Plazo de pago	2 días	7 días	79%

Fuente. Elaboración Propia

El modelo integrado prácticamente duplica o triplica el tiempo de anticipación de la línea base y logra emitir la mayoría de sus alertas antes del punto de no retorno. Esta ganancia es la que traduce la mejora predictiva en capacidad real de intervenir.

- **Accionabilidad: simulación de intervención**

Finalmente, una simulación por eventos discretos estimó cuántos estudiantes en riesgo serían efectivamente alcanzables bajo distintas políticas de disparo de la alerta. La Tabla 6 contrasta tres políticas: la de la línea base (disparo por umbral de riesgo académico), una política de disparo continuo y la política propuesta (disparo alineado con las ventanas administrativas).

Tabla 6. Accionabilidad estimada según política de disparo de alerta - valores relativos sobre población en riesgo

Política de disparo	En riesgo detectados	Alcanzables antes de compuerta	Deserciones evitables (est.)
Umbral académico (base)	100%	46%	12%
Disparo continuo	100%	58%	17%
Alineado a ventanas (propuesta)	100%	81%	26%

Fuente. Elaboración Propia

La política alineada con las ventanas administrativas eleva la proporción de estudiantes alcanzables antes de la compuerta del 46% al 81%, y aproximadamente duplica las deserciones potencialmente evitables frente a la línea base. Este resultado respalda la tercera proposición: sincronizar la alerta con la estructura administrativa incrementa la accionabilidad, con independencia de la exactitud del clasificador.

Discusión

Los resultados convergen en una idea: tratar las reglas administrativas como variables del modelo aporta valor en dos planos complementarios. En el plano predictivo, mejora las métricas sensibles a la clase deserción (PR-AUC y recall), porque codifica información sobre el momento y las opciones del estudiante que las variables académicas no capturan. En el plano operativo, y de manera acaso más importante, mejora la accionabilidad, al anticipar la alerta antes del punto de no retorno y permitir intervenciones que de otro modo llegarían tarde.

La implicación para la gestión institucional es directa. Un sistema de permanencia no debería limitarse a ordenar estudiantes por riesgo, sino a emitir la señal en el momento en que aún existe una maniobra administrativa viable. Esto sugiere rediseñar los protocolos de intervención para acoplarlos al calendario de compuertas, y no solo a los cortes académicos. También abre la puerta a diseñar las propias reglas administrativas (por ejemplo, la duración o ubicación de las ventanas) con criterios de permanencia.

Conclusiones

El estudio muestra que reencuadrar la alerta temprana de deserción como un problema de ingeniería de procesos, e integrar las reglas administrativas institucionales como variables explícitas, mejora tanto la capacidad predictiva como la accionabilidad del modelo. El riesgo de deserción se concentra alrededor de compuertas administrativas, las variables administrativas figuran entre las más influyentes, y alinear el disparo de la alerta con las ventanas administrativas incrementa de forma sustancial la proporción de estudiantes alcanzables antes del punto de no retorno.

Entre las limitaciones se cuentan la dependencia del calendario específico de la institución, con el consiguiente riesgo de sobreajuste institucional, la exigencia de un registro de eventos completo y de calidad, y la necesidad de validar la transferencia del modelo a más de una institución. El tratamiento de datos personales debe ajustarse a la normativa colombiana de protección de datos (Ley 1581 de 2012), con anonimización y aval del comité de ética. El trabajo futuro contempla la validación multiinstitucional y la optimización conjunta del clasificador y de la política de disparo.

Referencias

- Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education*, 12(2), 155 a 187.
- Ministerio de Educación Nacional de Colombia. (s. f.). Sistema para la Prevención de la Deserción en las Instituciones de Educación Superior (SPADIES).
- Spady, W. G. (1970). Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange*, 1(1), 64 a 85.
- Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2.^a ed.). University of Chicago Press.
- van der Aalst, W. (2016). *Process mining: Data science in action* (2.^a ed.). Springer.
- Congreso de la República de Colombia. (2012). Ley Estatutaria 1581 de 2012, sobre protección de datos personales.