

Recibido 2 Nov 2025

ReCIBE, Año 15 No. 1, mayo 2026

Aceptado 22 Abri 2026

Técnicas de Análisis de Supervivencia para la Identificación de Deserción en Estudiantes: Una Revisión Sistemática de la Literatura

Survival Analysis Techniques for Identifying Student Dropout: A Systematic Literature Review

Franz Jesús Rivera-Alcántara ¹

franz.j.rivera@outlook.com

Ángel J. Sánchez-García ¹

angesanchez@uv.mx

José Juan Muñoz-León ¹

juanmunoz@uv.mx

Jorge Octavio Ocharán-Hernández ¹

jocharan@uv.mx

José Luis Sánchez-Leyva ²

luissanchez01@uv.mx

¹ *Facultad de Estadística e Informática,
Universidad Veracruzana, Xalapa-Enríquez, Veracruz, México.*

² *Facultad de Contaduría y Administración,
Universidad Veracruzana, Coatzacoalcos, Veracruz, México.*

Resumen

Este trabajo presenta una Revisión Sistemática de la Literatura (RSL) sobre el uso de técnicas de Análisis de Supervivencia (AS) para predecir la deserción estudiantil para el problema de Ingeniería de Software. La educación en Ingeniería de Software enfrenta altas tasas de abandono, lo que resalta la necesidad de enfoques predictivos que apoyen la retención escolar. El AS, aplicado con éxito en áreas como la salud, permite modelar tanto la ocurrencia como el momento de un evento, ofreciendo ventajas frente a otros métodos. Siguiendo las directrices de Kitchenham y el enfoque Quasi-Gold Standard, se identificaron 36 estudios primarios publicados entre 2013 y 2025. Los resultados muestran que el modelo de riesgos proporcionales de Cox es la técnica más utilizada, junto con Kaplan-Meier y otros enfoques de regresión. Asimismo, se identificaron 29 variables relevantes y 16 métricas de validación. La revisión evidencia que el AS facilita la detección temprana de estudiantes en riesgo, aunque enfrenta retos como la necesidad de datos extensos y su limitada adopción en el ámbito educativo.

Palabras Clave: Educación en Ingeniería de Software; Análisis de Supervivencia; Deserción estudiantil; Revisión Sistemática de la Literatura.

Abstract

This work presents a Systematic Literature Review (SLR) on the use of Survival Analysis (SA) techniques to predict student dropout in the context of Software Engineering. Software Engineering education faces high attrition rates, highlighting the need for predictive approaches that support student retention. SA, successfully applied in fields such as healthcare, enables modeling both the occurrence and timing of an event, offering advantages over other methods. Following Kitchenham's guidelines and the Quasi-Gold Standard approach, 36 primary studies published between 2013 and 2025 were identified. The results show that Cox proportional hazards model is the most widely used technique, along with Kaplan-Meier and other regression approaches. In addition, 29 relevant variables and 16 validation metrics were identified. The review demonstrates that SA facilitates the early detection of at-risk students, although it faces challenges such as the need for extensive datasets and its limited adoption in the educational domain.

Keywords: Software Engineering Education; Survival Analysis; Student dropout; Systematic Literature Review.

1. Introducción

La Ingeniería de Software (IS) es una disciplina que adopta enfoques de ingeniería para desarrollar software de calidad [1]. La IS brinda técnicas de apoyo a la especificación, diseño y evolución del software, garantizando su calidad [2]. Para tal efecto, el ingeniero de software adopta un enfoque sistemático y organizado de su trabajo para producir software de alta calidad [2].

La educación en la IS desempeña un papel esencial en la formación de los estudiantes sobre tecnologías, procesos y prácticas que aseguren la calidad en el desarrollo de software [3]. Sin embargo, el número de desarrolladores de software titulados en IS es pequeño en comparación con el número de personas que desarrollan o modifican software [4]. El estudio de la deserción escolar (DE) en la IS y campos relacionados con la informática ha recibido una mayor atención debido a la necesidad de profesionales en estas carreras [5].

El Análisis de Supervivencia (AS) surge como propuesta por parte de investigadores de las comunidades de aprendizaje automático y minería de datos para enfrentar esta problemática [6]. El AS permite identificar los factores que afectan un evento de interés

[7]. En el ámbito de la DE, el AS facilita la identificación de estudiantes en riesgo y de los factores que influyen en esta condición, mediante el análisis de los datos presentados en su perfil de ingreso (PI) y TE.

A pesar del éxito de los métodos de AS en ámbitos como la sanidad, la salud y en recientes estudios enfocados en la DE, la aplicación de estos métodos ha sido escasamente explorada en comparación con otros enfoques en este problema. Esta carencia en la aplicación del AS en la DE resalta la necesidad de realizar una Revisión Sistemática de la Literatura (RSL) centrada en las técnicas de AS utilizadas como colaboración para resolver el problema de la DE.

Este trabajo aborda dicha brecha mediante una Revisión Sistemática de la Literatura (RSL) que consolida y analiza 36 estudios primarios publicados entre 2013 y 2025, siguiendo las directrices de Kitchenham y el enfoque Quasi-Gold Standard (QGS). La revisión ofrece un mapeo estructurado de las técnicas de AS más utilizadas (destacando la prevalencia del modelo de Riesgos Proporcionales de Cox y el estimador Kaplan-Meier) junto con 29 variables relevantes y 16 métricas de validación. Asimismo, identifica los principales retos y oportunidades de investigación para mejorar la detección temprana de estudiantes en riesgo. Al sintetizar el estado del arte, este trabajo aporta una referencia fundamental para futuras investigaciones y el desarrollo de modelos predictivos orientados a fortalecer la retención estudiantil en la educación en Ingeniería de Software.

El resto de este artículo se organiza de la siguiente manera: la Sección 2 presenta los trabajos relacionados. La Sección 3 describe la metodología aplicada en este estudio. La Sección 4 reporta los resultados experimentales, mientras que la Sección 5 ofrece una discusión de los hallazgos. La Sección 6 expone las amenazas a la validez, y la Sección 7 presenta las conclusiones.

2. Trabajo relacionado

Uno de los primeros antecedentes del uso de Análisis de Supervivencia (AS) aplicado a la deserción escolar (DE) fue el trabajo de Plank y colaboradores, quienes emplearon el modelo de regresión de Cox para estudiar la relación entre la DE en educación secundaria y el plan de estudios. En particular, analizaron cómo la combinación de educación técnica (CTE) y cursos académicos influía en la probabilidad de abandono, concluyendo que la exposición a ambos tipos de formación podía fortalecer la motivación y el apego del estudiante a la escuela [8]. Posteriormente, Ameri et al. [6] propusieron un marco de predicción temprana de la DE en la Wayne State University, también basado en el modelo de Cox, incorporando variables demográficas, socioeconómicas y del trayecto escolar. Sus resultados evidenciaron la utilidad de información previa a la matrícula y de factores académicos como créditos aprobados/reprobados y promedio semestral, para identificar estudiantes en riesgo y orientar intervenciones oportunas [6].

En un estudio más reciente, Lisboa y colaboradores [9] analizaron la DE en estudiantes de Ciencias de la Computación de la Universidad de Brasilia (UnB) mediante un modelo de regresión Log-Normal. Este permitió estimar el número de semestres hasta la deserción o el último seguimiento, destacando la influencia de factores como la modalidad de ingreso y el desempeño académico en la probabilidad de abandono [9]. Estos estudios reflejan la utilidad del AS como técnica robusta para modelar y comprender los factores asociados a la deserción en distintos niveles educativos y disciplinas.

A pesar de la creciente preocupación por las altas tasas de deserción en los programas de Ingeniería de Software, la literatura carece de una síntesis sistemática sobre la aplicación de técnicas de AS para predecir la deserción estudiantil en este dominio. Aunque el AS ha sido empleado con éxito en áreas como la salud para modelar tanto la

ocurrencia como el momento de un evento, su adopción en contextos educativos (y particularmente en la educación en Ingeniería de Software) sigue siendo limitada y dispersa. Los estudios existentes suelen centrarse en métodos de clasificación o regresión tradicionales, sin considerar la dimensión temporal que el AS permite analizar. Además, no existe una revisión integral que identifique qué técnicas, variables y métricas de validación han resultado más efectivas en la predicción de la deserción estudiantil.

En conjunto, la evidencia muestra la preocupación de las instituciones por fortalecer la retención escolar (RE) y alcanzar una mayor eficiencia terminal (ET). La deserción, particularmente en áreas de informática, sigue mostrando tasas elevadas pese a la creciente demanda de profesionales en Ingeniería de Software. Ante la insuficiencia de estrategias tradicionales, la aplicación del AS se presenta como una alternativa prometedora para anticipar riesgos y diseñar intervenciones focalizadas. Sin embargo, la escasez de estudios que empleen esta técnica en el análisis de la DE en programas de IS, junto con los crecientes índices de abandono, justifican la pertinencia de continuar investigando en esta línea.

3. Método de investigación

La metodología seguida para esta RSL se basó en las directrices propuestas en [10], que define tres etapas principales: planificación, ejecución y presentación de los hallazgos. Durante la etapa de ejecución, se aplicó una estrategia de búsqueda basada en el enfoque Quasi-Gold Standard propuesta en [11].

3.1 Planeación

Preguntas de investigación Para realizar este estudio, se establecieron cinco preguntas de investigación (PI) mostradas en la Tabla 1 con su respectiva motivación. Estas preguntas sirvieron como guía para extraer aspectos clave del AS en la deserción escolar.

Pregunta de Investigación	Motivación
PI1. ¿Cuáles son las técnicas de AS utilizadas en la predicción de deserción de estudiantes?	Identificar las técnicas de AS más utilizadas en el ámbito de la DE en estudiantes.
PI2. ¿Cuáles son las variables utilizadas para modelar el AS en la predicción de deserción de estudiantes?	Identificar los factores más utilizados para modelar el AS en el ámbito de la DE en estudiantes.
PI3. ¿Cuáles son las métricas para validar los modelos de AS en la predicción de deserción de estudiantes?	Identificar las métricas utilizadas para validar los modelos de AS en el ámbito de la DE en estudiantes.
PI4. ¿Cuáles son las ventajas de las técnicas de AS reportadas?	Identificar los beneficios de las técnicas de AS reportadas en la literatura en el ámbito de la DE en estudiantes.
PI5. ¿Cuáles son los retos de las técnicas de AS reportadas?	Identificar los desafíos de las técnicas de AS reportadas en la literatura en el ámbito de la DE en estudiantes.

Tabla 1. Preguntas de investigación y sus motivaciones.

Proceso de búsqueda La estrategia de búsqueda siguió el proceso manual propuesto por Zhang [11] que determinó las bibliotecas digitales más utilizadas en las RSL para la búsqueda de estudios relevantes en IS. Los recursos editoriales clave presentados en el estudio son IEEEExplore, ACM Digital Library, ScienceDirect (Elsevier) y SpringerLink.

Además, se consideraron las plataformas Wiley Online Library y Taylor and Francis Online, debido a la naturaleza multidisciplinaria de esta RSL. Estas fuentes proveen un lugar de búsqueda confiable para la búsqueda de estudios relevantes en el ámbito del AS para la predicción de la DE.

Concepto en idioma español	Concepto en idioma inglés	Términos y sinónimos alternativos
Análisis de Supervivencia	Survival Analysis	Survival Function, Survival Model, Survival Probability, Survival Time
Deserción Estudiantil	Student Dropout	School Dropout
Retención de Estudiantes	Student Retention	—

Tabla 2. Términos de búsqueda (palabras clave).

La cadena de búsqueda se construyó basándose en la propuesta realizada por [11]. Esta propuesta comprende la combinación de un método subjetivo y un método objetivo. Esta combinación permitió la identificación de los términos de búsqueda presentados en la Tabla 2, con el propósito de construir las cadenas de búsqueda presentadas en la Tabla 3, para la búsqueda automatizada de estudios primarios relacionados con el AS aplicado en la predicción de la DE.

ID	CB	ERS	ERO	EOCB	Recall	Effort
CB1	“survival*” AND (“school dropout” OR “student dropout” OR “student retention”)	6	5	1,366	83.33%	0.36%
CB2	“survival analysis” AND (“school dropout” OR “student dropout” OR “student retention”)	6	5	102	83.33%	4.90%
CB3	(“survival analysis” OR “survival function” OR “survival model” OR “survival probability” OR “survival time”) AND (“school dropout” OR “student dropout” OR “student retention”)	6	5	132	83.33%	3.78%

Tabla 3. Cadenas de búsqueda (ID: Identificador de la Cadena de Búsqueda; CB: Cadena de Búsqueda; ERS: Número de estudios relevantes seleccionados; ERO: Total de estudios relevantes obtenidos; EOCB: Total de estudios obtenidos por la cadena de búsqueda).

Para la evaluación de la cadena de búsqueda, se utilizó la técnica propuesta por [11]. Esta técnica utiliza un enfoque objetivo, que emplea un conjunto de criterios cuantitativos para valorar el rendimiento de la estrategia de búsqueda: sensibilidad (recall) y precisión (effort). La sensibilidad se define como la proporción de estudios relevantes recuperados para un tema determinado; y la precisión es la proporción de estudios recuperados que son estudios relevantes. La sensibilidad y la precisión correspondientes a la cadena de búsqueda está dada por (1) y (2):

$$(1) \text{ Recall} = \frac{\text{Retrieved relevant studies}}{\text{Relevant studies}}$$

$$(2) \text{ Precision} = \frac{\text{Retrieved relevant studies}}{\text{Retrieved studies}}$$

La evaluación de la cadena de búsqueda se realizó en el motor de búsqueda multidisciplinario SpringerLink, la selección del motor de búsqueda se realizó. Esta evaluación revela la competencia de las cadenas de búsqueda sugeridas para la recuperación de estudios pertinentes, basándose en los valores adquiridos para los criterios de sensibilidad (recall) y precisión (effort). Esta apreciación evidencia que las cadenas de búsqueda propuestas alcanzaron un porcentaje de 83.33% con respecto al total de ERS.

En el proceso de evaluación propuesto por [11], se sugiere un umbral en el criterio recall para la aprobación de la cadena de búsqueda, que son los valores que superan el 80%. En esta propuesta, el valor de recall adquiere mayor importancia al representar el valor de estudios relevantes obtenidos por la cadena de búsqueda. Por este motivo, la cadena de búsqueda seleccionada para esta investigación es la cadena de búsqueda CB3. Esta cadena de búsqueda demostró una efectiva recuperación del total de ERS, agregando los términos relacionados pertinentes para la investigación. Aunque la cadena de búsqueda CB2 presentó un incremento en el valor correspondiente al criterio effort en relación con la cadena de búsqueda CB3, se consideró la cadena de búsqueda CB3 debido a su adaptabilidad en los demás motores de búsqueda seleccionados.

("survival analysis" OR "survival function" OR "survival model" OR "survival probability" OR "survival time") AND ("school dropout" OR "student dropout" OR "student retention")

Extracción de datos El objetivo de esta subsección del proceso de revisión es obtener los datos necesarios de los estudios primarios para responder a las preguntas de la investigación [10]. La plantilla representada en la Tabla 4 expone los datos generales que se extraerán de los estudios primarios seleccionados.

Criterios de selección Los criterios de selección de estudios primarios se formulan con el fin de identificar aquellos estudios capaces de aportar pruebas que sean de relevancia para las preguntas de investigación [10]. Estos criterios se expresan en dos conjuntos: criterios de inclusión y criterios de exclusión mostrados en las Tablas 5 y 6 respectivamente.

Dato	Descripción
ID	Identificador único del estudio generado con la nomenclatura EP-#, que corresponde a Estudio Primario (EP) - #número.
Título	Título del estudio.
Autores	Nombres de los autores responsables de la publicación del estudio.
Año de publicación	Año de publicación del estudio primario, comprendido entre los años 2013–2025, alineándose a los criterios de inclusión.
Fuente	Dirección URL donde se encuentra el estudio en línea.

Tipo de publicación	Tipo de publicación (conferencia o journal).
Referencia o detalles de la publicación	Referencia de la publicación.
Palabras clave	Palabras clave establecidas en el estudio.
Venue	Nombre de la conferencia o journal donde fue presentado el estudio.
DOI	Identificador único del estudio publicado.
Abstract o resumen	Resumen (abstract) establecido en el estudio.
Pregunta(s) de investigación relacionada(s)	Preguntas de investigación que el estudio responde.
Técnica de AS utilizada en la predicción de DE	Técnica de Aprendizaje Supervisado (AS) utilizada en la predicción de deserción de estudiantes. (PI1).
Variable(s) utilizadas para modelar el AS en la predicción de DE	Variabes utilizadas para modelar el AS en la predicción de deserción de estudiantes. (PI2).
Métrica utilizada para validar el modelo de AS en la predicción de DE	Métrica utilizada para validar el modelo de AS en la predicción de deserción de estudiantes. (PI3).
Ventaja(s) de la técnica de AS en la predicción de DE	Ventaja(s) reportadas por la técnica de AS utilizada en la predicción de deserción de estudiantes. (PI4).
Reto(s) de la técnica de AS en la predicción de DE	Reto(s) reportados por la técnica de AS utilizada en la predicción de deserción de estudiantes. (PI5).

Tabla 4. Formato de extracción de los datos.

ID	Criterio
CI1	El estudio está publicado entre 2013 y 2025.
CI2	El estudio está escrito en el idioma inglés.
CI3	Se tiene acceso completo al estudio.
CI4	Al leer el título y el resumen (abstract) del estudio se tienen indicios de responder al menos una pregunta de investigación.
CI5	Al leer completamente el estudio se responde al menos una pregunta de investigación.

Tabla 5. Criterios de inclusión.

ID	Criterio
CE1	El estudio está publicado como diapositiva, resumen, reporte técnico o capítulo de libro.
CE2	Al leer el título y el resumen (abstract) del estudio se identifica que el enfoque principal se encuentra fuera del ámbito de la educación.
CE3	El estudio está duplicado.

Tabla 6. Criterios de exclusión.

3.2 Ejecución

Cuatro fases fueron definidas, cada una con criterios de inclusión y exclusión específicos mostrados a continuación:

Fase 1: Se incluyeron estudios publicados entre 2013 y julio de 2025 y además escritos en el idioma inglés.

Fase 2: Se incluyeron estudios con acceso completo y se excluyeron aquellos que estuvieran publicados como presentación, resumen, reporte técnico o capítulo de libro.

Fase 3: Se incluyeron aquellos estudios que al leer el título y el resumen (abstract) del estudio se tienen indicios de responder al menos una pregunta de investigación.

Fase 4: Finalmente se incluyeron aquellos estudios que contestaran al menos una pregunta de investigación y se excluyeron los estudios duplicados.

3.3 Reporte

Se seleccionó utilizar una síntesis narrativa siguiendo los pasos presentados por [12]. En esta síntesis, se analizó los hallazgos, identificando patrones en las gráficas y tabulaciones.

4. Resultados

Como resultado de la estrategia de búsqueda y el proceso de selección, se identificaron 36 estudios primarios que respondieron a las preguntas de investigación. La síntesis narrativa se desarrolló siguiendo la metodología propuesta por Popay [12].

La fuente Taylor & Francis fue la fuente que más aportó a los estudios primarios con el 27.7% (10 estudios) seguido de Springer Link con 25% (9 estudios) y Wiley 19.4% (7 estudios). ACM DL y IEEE Xplore aportaron la misma cantidad de estudios con un 11.1% (4 estudios), seguida Science Direct con 5.55% (2 estudios). Los 36 estudios primarios se pueden encontrar en [13].

4.1 P11. ¿Cuáles son las técnicas de Análisis de Supervivencia utilizadas en la predicción de deserción de estudiantes?

A partir de la Fig. 1, se puede observar que el modelo Cox proportional hazards model fue la técnica de AS más utilizada, con un 48.57% del total de estudios primarios que respondieron la pregunta de investigación. La segunda técnica más utilizada es el modelo de Kaplan-Meier, con un 31.42% del total de estudios primarios que respondieron la pregunta de investigación. El modelo Logistic regression model y el enfoque Discrete-time hazard model ocuparon el tercer lugar, con un 11.42% del total de estudios primarios que respondieron la pregunta de investigación, respectivamente. El modelo Multinomial logistic regression model fue la última técnica en reportar una frecuencia mayor en comparación con el resto de las técnicas que reportaron una única intervención en los

estudios primarios, con un 8.57% del total de estudios primarios que respondieron la pregunta de investigación.

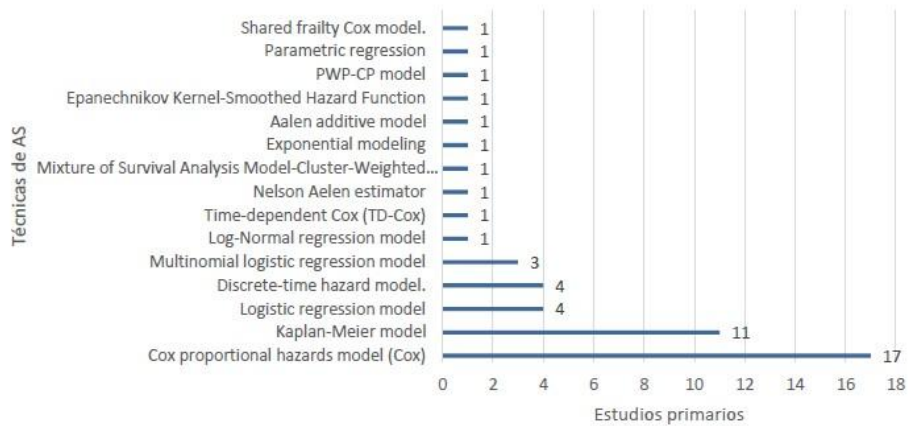


Figura 1. Distribución de estudios primarios por técnica de AS.

La razón de la preferencia en la adopción del modelo Cox proportional hazards model en el ámbito educativo es su amplio historial de uso en otros ámbitos, como lo determina [6] en su intervención en la Wayne State University, en donde afirma que, a pesar del éxito de los métodos de AS en áreas como la salud, la ingeniería, etc., solo se ha intentado utilizar estos métodos de forma limitada en el problema de la retención de estudiantes.

4.2 PI2. ¿Cuáles son las variables utilizadas para modelar el Análisis de Supervivencia en la predicción de deserción de estudiantes?

Se extrajeron un total de 29 variables clave pertenecientes a entornos universitarios, cursos de plataformas MOOC y entorno social particular. Estos atributos caracterizaron a la población estudiantil en ámbitos como los datos demográficos, atributos financieros, antecedentes académicos e información semestral.

La Figura 2 muestra las variables clave en el contexto universitario, en donde se aprecia que el género, el promedio, el puntaje de examen de admisión, la educación de los padres, la edad, el promedio de escuela secundaria o bachillerato, las bajas por semestre, el abandono de un curso y la ubicación de residencia son los factores que más influyeron en el abandono de los estudiantes de institutos de educación superior.

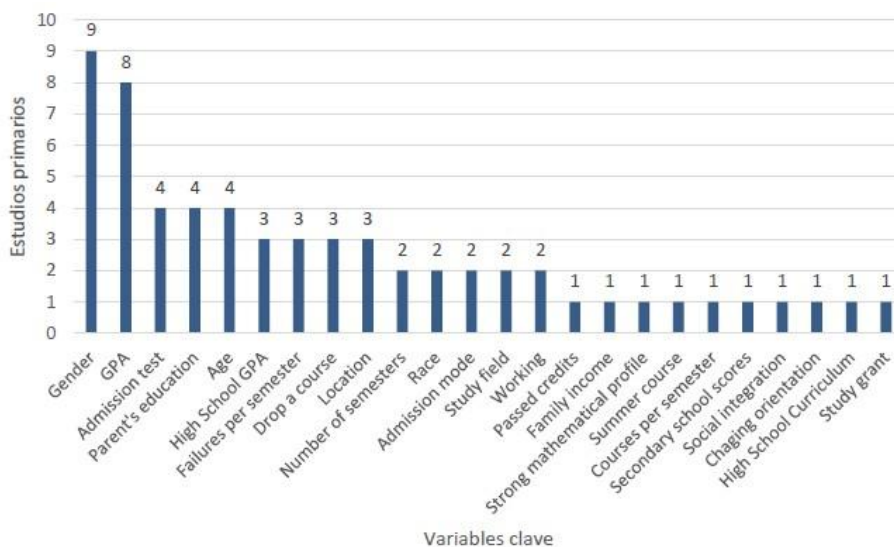


Figura 2. Variables clave en el contexto universitario.

4.3 PI3. ¿Cuáles son las métricas para validar los modelos de Análisis de Supervivencia en la predicción de deserción de estudiantes?

Del total de estudios primarios que respondieron esta pregunta de investigación, se extrajeron 16 métricas de validación de modelos de AS en la predicción de la DE. La Figura 3 muestra la distribución de estudios primarios segmentados por métrica reportada.

Las métricas más frecuentes proporcionan una visión detallada de cómo estos modelos funcionan específicamente en el contexto educativo. Entre las métricas reportadas, el Log-rank test es el más común con un 30.76% de los estudios, seguido de cerca por Recall, AUC, Akaike Information Criterion (AIC) y Concordance Index (C-index), cada uno con un 23.07%. Además, métricas como Accuracy, Precision, F-measure, y Bayesian Information Criterion (BIC) muestran una frecuencia del 15.38%.

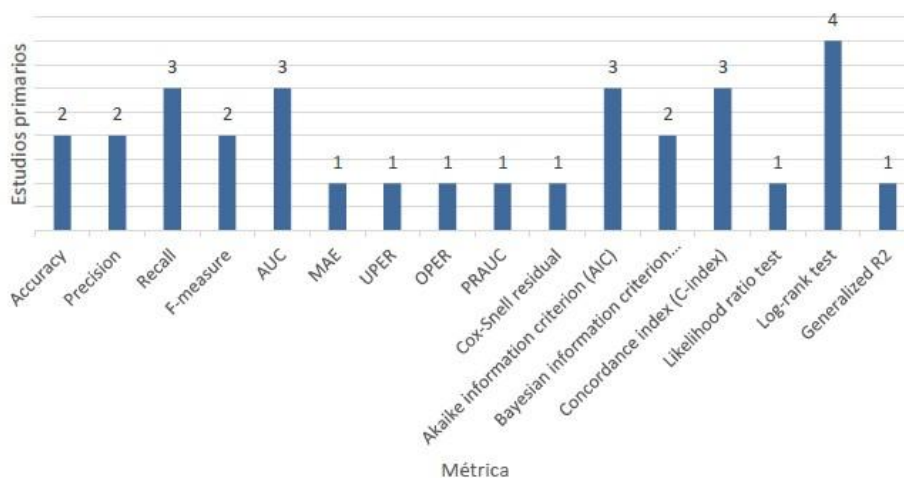


Figura 3. Distribución de estudios primarios por métrica reportada.

4.4 PI4. ¿Cuáles son las ventajas de las técnicas de Análisis de Supervivencia reportadas?

En la literatura educativa, los investigadores resaltan los beneficios de aplicar técnicas basadas en AS para obtener información valiosa sobre los estudiantes en riesgo. A continuación, la Tabla 7 presenta las ventajas generales del AS identificadas durante la revisión.

4.5 PI5. ¿Cuáles son los retos de las técnicas de Análisis de Supervivencia reportadas?

De los 36 estudios primarios, 3 reportaron retos de las técnicas de AS utilizadas para la predicción de deserción de estudiantes. Asimismo, se identificaron 2 retos generales del AS aplicado en el ámbito de la DE, proporcionados por los estudios EP-1 y EP-23. Tras el análisis de los estudios primarios, se encontraron un total de 6 retos específicos de las técnicas de AS, los cuales se relacionan únicamente con 3 modelos de AS del total de 15 modelos identificados.

El primer reto general adquirido de los estudios primarios es el uso limitado del AS en el ámbito de la retención de estudiantes y de la DE. Existe una tendencia por parte de los autores en la modificación y extensión de los modelos estadísticos de AS para mejorar el ajuste a los datos. A pesar de que el modelo de AS tiene más flexibilidad para tratar el problema de la retención de estudiantes, hay pocos esfuerzos en la literatura en este ámbito de la educación.

Otro reto más específico del AS y en comparación con otros métodos de estimación utilizados en la educación, es que tiende a requerir más información, esto puede suponer un reto importante para aquellos estudios con limitada cantidad de datos disponibles. Este reto se presenta en el estudio realizado por [14], en donde se demuestra que el AS necesitó una mayor cantidad de datos en comparación con los modelos logit y probit.

Ventaja	Estudio
El AS puede incorporar dos componentes: si se produce o no un acontecimiento (en el ámbito de la DE sería el abandono) y cuándo se producirá el acontecimiento. Por lo tanto, la ventaja de utilizar el AS en lugar de regresión logística u otros métodos de minería de datos es la capacidad de añadir el componente temporal al modelo.	EP-1
En el AS, cualquier cambio en el riesgo de abandono cuando cambia una de las características del estudiante es fácil de observar.	EP-11
El AS no requiere grandes muestras de entrenamiento para lograr un aprendizaje automático que clasifique a un estudiante en un riesgo específico de abandono.	EP-11
En el AS, los datos censurados hacen referencia a datos de estudiantes que no presentaron el evento de interés (abandono) durante el período de observación. Barragán y colaboradores resaltan el beneficio del AS al incorporar información censurada.	EP-11
El AS permite evaluar cómo diferentes grupos dentro de una población difieren respecto al tiempo que toma hasta que ocurre el evento estudiado (en el ámbito educativo, el abandono). Estas diferencias pueden ser por género, área de conocimiento, y demás atributos que caractericen la población.	EP-11
El AS puede identificar cuales variables o factores tienen un impacto significativo en la probabilidad de abandonar.	EP-11

Tabla 7. Ventajas del uso de AS en la deserción escolar.

5. Discusión

Los hallazgos de esta revisión y la aplicación de estos modelos encontrados en la literatura permitirían a las autoridades educativas:

- Identificar a los estudiantes que requieren intervención temprana, brindando así apoyo oportuno antes de que se materialice el riesgo de deserción escolar.
- Implementar programas de tutoría y mentoría en colaboración con el profesorado (docentes y estudiantes).
- Reconocer a los estudiantes con buenos candidatos para becas u oportunidades de ayuda financiera.
- Revisar y adaptar el currículo académico para alinearlo mejor con las necesidades y trayectorias de aprendizaje de los estudiantes.
- Monitorear los factores no académicos que influyen en el rendimiento académico, incluyendo el nivel socioeconómico, las características sociodemográficas, las condiciones laborales, los aspectos psicológicos y otras variables contextuales.

6. Amenazas a la validez

En este estudio se consideraron únicamente aquellos motores de búsqueda de acceso disponible por medio de la institución. Sin embargo, durante el proceso de la investigación no se contó con el acceso completo a algunos motores de búsqueda, esta

limitación en la disponibilidad de algunos estudios revela la posible omisión de algún estudio relevante en la revisión.

Durante el proceso de creación de las cadenas de búsqueda consideradas para conducir la revisión, se evaluaron utilizando un Quasi-Gold Standard conformado únicamente por estudios relevantes pertenecientes al motor de búsqueda SpringerLink debido al enfoque multidisciplinario del tema de investigación. A pesar de estas limitaciones, se presentan las siguientes medidas para mitigar las posibles consecuencias de los elementos descritos anteriormente:

- Pese a la falta de acceso completo en algunas bases de datos, la selección de los motores de búsqueda se basó en el estudio realizado [11], que considera las fuentes de información más utilizadas en las revisiones sistemáticas e investigaciones relacionadas con la IS.

- Se siguió de forma sistematizada el proceso propuesto por [10] para planificar y conducir la revisión, el método para identificar estudios relevantes en IS publicado por Zhang y colaboradores [11] y se sintetizaron los resultados reportados por la literatura utilizando elementos de la síntesis narrativa diseñada por [12].

- La cadena de búsqueda CB3 se seleccionó debido a los valores reportados para los criterios Recall y Effort durante su evaluación en el motor de búsqueda SpringerLink, que cumplen con los valores aceptados en el método de evaluación propuesto por Zhang y colaboradores. Asimismo, la cadena de búsqueda se probó en cada fuente de información seleccionada para asegurar su adaptación.

- Durante las fases de planificación, conducción y documentación de la revisión propuestas por [10], se consultó la opinión de profesionales en el ámbito estadístico y de inteligencia artificial para asegurar la calidad de la revisión.

- Otra limitación importante es la ausencia de una evaluación formal de la calidad de los estudios primarios. Si bien se consideró que el proceso de revisión por pares de revistas y conferencias proporciona un nivel básico de rigor, dicho proceso no es homogéneo entre distintas sedes, y sus estándares pueden variar significativamente. Sin una evaluación sistemática de aspectos como la solidez metodológica, la validez de los conjuntos de datos, la reproducibilidad de los experimentos y la claridad en la presentación de los resultados, la robustez de las conclusiones extraídas de los estudios incluidos puede verse debilitada.

7. Conclusión

Esta investigación reporta el proceso y resultados de una Revisión sistemática de la Literatura, con el objetivo de conocer el estado actual de las técnicas de AS utilizadas en el ámbito de la predicción de estudiantes en riesgo.

Las respuestas a las preguntas de investigación revelaron la utilidad del AS en el contexto educativo. Durante la revisión se identificaron 36 estudios primarios, que reportaron un total de 15 técnicas de AS, 29 variables clave, 16 métricas y detallaron las ventajas de la aplicación del AS en la predicción de estudiantes en riesgo. Además, se identificó un área de oportunidad en la especificación de los retos que rodean al AS. La literatura ha demostrado interés en la comprobación del éxito del AS como herramienta de prevención para el problema de la DE. Sin embargo, existe una amplia área de oportunidad en la identificación de las limitaciones relacionadas a su aplicación.

Este estudio permitirá conocer el estado del arte de las técnicas de AS utilizadas en el ámbito de la DE. Brindando información sobre los propósitos de las técnicas identificadas, las ventajas y retos inherentes a estos métodos y al AS en general. Los estudios reportan la efectividad de su uso en un contexto real, beneficiando al personal de las instituciones a implantar estrategias de retención efectivas.

El trabajo futuro se centra en la recopilación de más datos que caractericen a los estudiantes, con el objetivo de generar modelos más robustos y de mayor capacidad predictiva. Asimismo, existe un área de oportunidad en la identificación de los retos inherentes en la aplicación de técnicas de AS en la DE.

Referencias

- Y. Wang, *Software engineering foundations: A software science perspective*, Auerbach Publications, 2007. Sommerville, *Software engineering tenth edition* (2016).
- Cico, *Lean software startup practices and software engineering education*, in: *Proceedings of the ACM/IEEE 44th International Conference on Software Engineering: Companion Proceedings, 2022*, pp. 281–285.
- N. R. Mead, D. Garlan, M. Shaw, *Half a century of software engineering education: The cmu exemplar*, *IEEE Software* 35 (5) (2018) 25–31.
- S. Bayona-Oré, *Dropout in computer science, systems engineering and software engineering programs*, in: *World Conference on Information Systems and Technologies*, Springer, 2023, pp. 592–599.
- S. Ameri, M. J. Fard, R. B. Chinnam, C. K. Reddy, *Survival analysis based framework for early prediction of student dropouts*, in: *Proceedings of the 25th ACM international on conference on information and knowledge management*, 2016, pp. 903–912.
- H. Shi, Y. Zhou, *Stay or leave? exploring student factors associated with dropout patterns in massive open online courses*, in: *2023 IEEE International Conference on Advanced Learning Technologies (ICALT)*, IEEE, 2023, pp. 26–30.
- S. B. Plank, S. DeLuca, A. Estacion, *High school dropout and the role of career and technical education: A survival analysis of surviving high school*, *Sociology of Education* 81 (4) (2008) 345–370.
- J. B. F. Gomes, M. Holanda, C. C. Koike, M. T. L. Costa, et al., *Study on computer science undergraduate students dropout at the university of brasilia*, in: *2023 IEEE Frontiers in Education Conference (FIE)*, IEEE, 2023, pp. 1–7.
- B. A. Kitchenham, D. Budgen, P. Brereton, *Evidence-Based Software Engineering and Systematic Reviews*, Chapman and Hall/CRC, 2015. doi:10.1201/b19467.
- H. Zhang, M. A. Babar, P. Tell, *Identifying relevant studies in software engineering*, *Information and Software Technology* 53 (2011) 625–637. doi:10.1016/j.infsof.2010.12.010.
- J. Popay, L. Arai, M. Rodgers, N. Britten, *Guidance on the conduct of narrative synthesis in systematic reviews: A product from the esrc methods programme* (2006). doi:10.13140/2.1.1018.4643.
- J. Sanchez-Garcia, *Apéndice a: Repositorio estudios primarios* (2025). doi: 10.5281/zenodo.17518840. URL <https://doi.org/10.5281/zenodo.17518840>
- K. Boualaphet, H. Goto, *Determinants of school dropout in lao people’s democratic republic: A survival analysis*, *Journal of International Development* 32 (6) (2020) 961–975.

NOTAS BIOGRÁFICAS



Franz Jesús Rivera-Alcántara, Licenciado en Ingeniería de Software por la Universidad Veracruzana, en la Facultad de Estadística e Informática. En su trayectoria académica, ha mostrado especial interés en aplicaciones de inteligencia artificial y ciencia de datos. Colaboró en el desarrollo de una solución basada en Retrieval-Augmented Generation (RAG) en Sophinauta LTD.



Ángel Juan Sánchez-García es Licenciado en Informática, Maestro en Inteligencia Artificial, Especialista en Métodos Estadísticos y Doctor en Inteligencia Artificial por la Universidad Veracruzana, México. Actualmente es profesor en la Facultad de Estadística e Informática de la Universidad Veracruzana. Es miembro del Cuerpo académico Ingeniería y Tecnología de Software. Es miembro desde 2018 del Sistema Nacional de Investigadoras e Investigadores del CONAHCYT (actualmente Nivel 1, área 8) y cuenta con el reconocimiento del Programa de Desarrollo Profesional Docente (PRODEP) desde 2018. Su trabajo de investigación incluye las áreas de Aprendizaje automático e Inteligencia Artificial aplicada a la Ingeniería de software. Página web: www.uv.mx/personal/angesanchez.



José Juan Muñoz-León es licenciado en Matemáticas y Maestro en Gestión de Calidad por la Universidad Veracruzana. Además, es Maestro y Doctor en Didáctica de las Matemáticas y las Ciencias Experimentales por la Universidad Autónoma de Barcelona. Actualmente es profesor en la Facultad de Estadística e Informática de la Universidad Veracruzana. Responsable del grupo de investigación Data Science Applications Reserch Group y con reconocimiento del Programa de Desarrollo Profesional Docente (PRODEP) hasta 2028. Su trabajo de investigación incluye áreas de analítica de datos y educación en el nivel superior.



Jorge Octavio Ocharán-Hernández obtuvo su Maestría en Ingeniería de Software y su Doctorado en Ciencias de la Computación en la Universidad Veracruzana, Veracruz, México. En Ingeniero en Tecnologías Estratégicas de la Información por la Universidad Anáhuac Veracruz. Actualmente es profesor de tiempo completo en la Facultad de Estadística e Informática de la Universidad Veracruzana. Cuenta con más de 20 años de experiencia en la industria y la formación en Ingeniería de Software. Sus intereses de investigación incluyen la ingeniería de requisitos, la arquitectura de software, el diseño de software y API, y la

aplicación de la inteligencia artificial en la ingeniería de software. Página web: <https://uv.mx/personal/jocharan/>.



El Dr. José Luis Sánchez-Leyva es Director General de Desarrollo Académico e Innovación Educativa de la Universidad Veracruzana, Académico de Carrera de Tiempo Completo Titular C con adscripción en la Facultad de Contaduría y Administración de la Universidad Veracruzana en la Región Coatzacoalcos-Minatitlán; es miembro del Sistema Nacional de Investigadores Nivel 1y Miembro del Padrón Veracruzano de Investigadores del Consejo Veracruzano de Investigación Científica y Desarrollo Tecnológico; cuenta con reconocimiento de Perfil Deseable del Programa para el Desarrollo Profesional Docente (PRODEP) y con la Certificación Profesional como Licenciado en Administración. En su formación profesional, es Doctor en Gobierno y Administración Pública; Maestro en Gobierno y Asuntos Públicos y Licenciado en Administración de Empresas. Además, es Responsable del Cuerpo Académico en Consolidación (UV-CA-506) “Economía del Conocimiento e Innovación” y fungió como director general del Área Académica Económico-Administrativa de 2022 a 2025 y como director de la Facultad de Contaduría y Administración en la Región Coatzacoalcos-Minatitlán de 2013 a 2018. El Dr. Sánchez Leyva es autor y coautor de artículos publicados en revistas y capítulos de libros, participa como conferencista y ponente en congresos nacionales e internacionales y es miembro de comité editoriales en diversas revistas y congresos.



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 2.5 México.