

Estilos de aprendizaje y abandono universitario: un enfoque metacognitivo y de analítica predictiva

GABRIELLA GIULIA PULCINI¹

Università di Camerino - Italia

Recibido el 15-07-26; primera evaluación el 05-08-26;
aceptado el 25-08-26

RESUMEN

El abandono universitario es un fenómeno multifactorial relacionado con variables académicas, personales y de aprendizaje. Este estudio analiza la relación entre estilos de aprendizaje, metacognición y abandono universitario en un curso *e-learning* de una universidad italiana. Se realizó un estudio exploratorio con 257 estudiantes y 22 variables predictoras procedentes del cuestionario CAMEA40, la trayectoria educativa, características personales y registros académicos. Se entrenaron seis modelos de clasificación: regresión logística, SVM, K-vecinos más cercanos, Random Forest, Naive Bayes y perceptrón multicapa. El desempeño se evaluó mediante exactitud, precisión, sensibilidad y F1-score. Random Forest obtuvo los mejores resultados: exactitud 0,91, precisión 0,91, sensibilidad 0,83 y F1-score 0,86. Los hallazgos, aunque requieren validación con muestras más amplias, sugieren que la analítica predictiva puede apoyar la detección temprana y la tutoría.

Palabras clave: Estilos de aprendizaje; metacognición; abandono universitario; analítica predictiva; aprendizaje automático.

¹ Doctora en Química por la Universidad de Camerino (Italia) y cuenta con formación en asesoría educativa, tutoría online y formación de formadores. Su trabajo se centra en estilos de aprendizaje, orientación, tutoría, e-learning e innovación didáctica. Desde 2000 es profesora de química en bachillerato y desde 2015 imparte Química General e Inorgánica en la UNICAM. También colabora con los servicios de tutoría y orientación en estudios sobre abandono escolar. Participa en comités científicos internacionales y coordina desde 2016 el programa “Conócete a ti mismo. Descubre tu estilo de aprendizaje”. Correo electrónico: gabriellagiulia.pulcini@unicam.it ORCID: <http://orcid.org/0000-0002-9044-6641>

Learning styles and university dropout: A metacognitive and predictive analytics approach

ABSTRACT

University dropout is a multifactorial phenomenon associated with academic, personal, and learning-related variables. This study analyzes the relationship between learning styles, metacognition, and university dropout within an e-learning course at an Italian university. An exploratory study was conducted with 257 students, utilizing 22 predictor variables derived from the CAMEA40 questionnaire, educational trajectories, personal characteristics, and academic records. Six classification models were trained: logistic regression, support vector machines (SVM), K-nearest neighbors, Random Forest, Naive Bayes, and multilayer perceptron. Model performance was evaluated using accuracy, precision, sensitivity (recall), and the F1-score. Random Forest achieved the best overall performance, with an accuracy of 0.91, precision of 0.91, sensitivity of 0.83, and an F1-score of 0.86. Although these findings require validation with broader samples, they suggest that predictive analytics can effectively support early detection and academic mentoring.

Keywords: Learning styles; metacognition; university dropout; predictive analytics; machine learning.

Estilos de aprendizagem e evasão universitária: uma abordagem metacognitiva e de análise preditiva

RESUMO

A evasão universitária é um fenômeno multifatorial associado a variáveis acadêmicas, pessoais e de aprendizagem. Este estudo analisa a relação entre estilos de aprendizagem, metacognição e evasão universitária em um curso na modalidade *e-learning* de uma universidade italiana. Foi realizado um estudo exploratório com 257 estudantes e 22 variáveis preditoras derivadas do questionário CAMEA40, do percurso educacional, de características pessoais e de registros acadêmicos. Foram treinados seis modelos de classificação: regressão logística, máquinas de vetores de suporte (SVM), K-vizinhos mais próximos, Random Forest, Naive Bayes e perceptron multicamadas. O desempenho foi avaliado por meio da acurácia, precisão, sensibilidade (*recall*) e F1-score. O modelo Random Forest obteve os melhores resultados: acurácia de 0,91, precisão de 0,91, sensibilidade de 0,83 e F1-score de 0,86. Os resultados indicam que a análise preditiva pode auxiliar na detecção precoce e na tutoria, embora necessitem de validação com amostras mais amplas.

Palavras-chave: Estilos de aprendizagem; metacognição; evasão universitária; análise preditiva; aprendizado de máquina.

1. INTRODUCCIÓN

El abandono universitario constituye un fenómeno multifactorial en el que intervienen variables académicas, personales, sociodemográficas y relacionadas con la integración del estudiante en la institución. En Italia, el abandono universitario continúa siendo un problema relevante y se concentra especialmente en las primeras etapas de la trayectoria académica. Según la Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca (Anvur, 2026), en la cohorte de estudiantes matriculados en programas de grado en 2018 y 2019, el 26,4% había interrumpido sus estudios seis años después de la matrícula. Este dato refuerza la necesidad de fortalecer las estrategias de orientación y de identificar tempranamente las situaciones de vulnerabilidad académica.

En este contexto se desarrolla el curso *e-learning* “Conócete a ti mismo. Descubre tu estilo de aprendizaje”, dirigido principalmente a estudiantes de primero y segundo año de una universidad italiana. El curso se fundamenta en la necesidad de ofrecer al estudiantado espacios estructurados para reconocer sus características, comprender sus formas de aprender y reflexionar sobre sus recursos, dificultades, intereses y metas académicas. Su propuesta combina actividades de autoconocimiento, reflexión guiada, trabajo metacognitivo y orientación académica, organizados en un itinerario formativo.

La metacognición constituye uno de los fundamentos conceptuales del proyecto, puesto que permite al estudiante reconocer, supervisar y regular sus propios procesos de aprendizaje. Desde esta perspectiva, los estilos de aprendizaje no se emplean como etiquetas fijas ni como categorías determinantes, sino como recursos orientativos para promover la reflexión sobre las preferencias, los hábitos y las estrategias de aprendizaje.

Durante su desarrollo, el curso ha incorporado recursos digitales y herramientas de inteligencia artificial (IA) como medios de apoyo para la elaboración de actividades, la interpretación de resultados y la planificación de objetivos personales. Sin embargo, estas tecnologías no se consideran instrumentos autónomos de decisión. La literatura reciente señala que la inteligencia artificial y la analítica del aprendizaje pueden contribuir a la identificación temprana de estudiantes en situación de riesgo, siempre que se integren en marcos pedagógicos y éticos que contemplen la privacidad, la transparencia, la prevención de sesgos y la supervisión humana (Ovtšarenko, 2026; Pacheco et al., 2025).

En este estudio, la experiencia formativa del curso se articula con un análisis preliminar de datos educativos orientado a identificar variables asociadas con el abandono universitario y a comparar el desempeño de distintos modelos

de clasificación. De este modo, la analítica predictiva se concibe como una herramienta de apoyo a la orientación y a la tutoría, no como un mecanismo para etiquetar estudiantes ni para sustituir el juicio profesional. El propósito es generar información que permita diseñar intervenciones educativas oportunas y contextualizadas, especialmente durante los primeros años de la trayectoria universitaria.

El objetivo general de este estudio es analizar la relación entre los estilos de aprendizaje, la metacognición y el abandono universitario mediante un enfoque de analítica predictiva aplicado a los datos recopilados o generados a través del curso *e-learning*. De manera específica, se pretende: describir los fundamentos metodológicos del curso; caracterizar las variables académicas, personales y relacionadas con los estilos de aprendizaje incluidas en el conjunto de datos; entrenar y comparar seis modelos de clasificación (regresión logística, máquinas de vectores de soporte, K-vecinos más cercanos, Random Forest, Naive Bayes y perceptrón multicapa), y discutir las posibilidades y limitaciones del uso de estos modelos como apoyo a las estrategias de orientación y permanencia universitaria.

2. MARCO CONCEPTUAL

El marco conceptual de esta experiencia integra cuatro ejes principales: autoconocimiento, metacognición, estilos de aprendizaje y analítica del aprendizaje. Estos ejes no se conciben como compartimentos aislados, sino como dimensiones complementarias de una propuesta formativa orientada a favorecer la reflexión, la autorregulación y la permanencia universitaria.

El autoconocimiento constituye el punto de partida del itinerario pedagógico. Se entiende aquí como la posibilidad de que el estudiante reconozca sus recursos, dificultades, intereses, expectativas y metas académicas, no como una descripción fija de su personalidad, sino como un proceso reflexivo que ayuda a interpretar la propia trayectoria universitaria y a formular objetivos realistas de mejora. En contextos universitarios complejos, el autoconocimiento no es un ejercicio accesorio, sino una condición relevante para sostener procesos de aprendizaje más conscientes y autónomos.

La metacognición proporciona el soporte teórico de este enfoque. Entendida como el conocimiento y la regulación de los propios procesos cognitivos, esta le permite al estudiantado observar cómo aprende, además de identificar obstáculos y ajustar estrategias. La revisión sistemática de Zapata-Zapata et al. (2024) confirma la relevancia de los juicios metacognitivos en la educación superior y su vínculo con la autorregulación del aprendizaje, lo

que respalda la inclusión de actividades orientadas a la autoobservación y a la reflexión sobre el propio desempeño.

El tercer eje está constituido por los estilos de aprendizaje. El Cuestionario Honey-Alonso de Estilos de Aprendizaje (Chaea) (Alonso et al., 1994) sirve como referencia teórica de base para comprender las preferencias vinculadas con los estilos activo, reflexivo, teórico y pragmático. Sobre esta base, el CAMEA40 (Cuestionario Adaptado para Monitorizar Estilos de Aprendizaje) presentado por Madrigal Gil y Trujillo Torres (2017), se emplea en este estudio como un instrumento de apoyo al acompañamiento tutorial y a la reflexión metacognitiva. La versión italiana del CAMEA40 fue traducida, adaptada y evaluada previamente en contextos universitarios; esta documentación (Pulcini et. al., 2018) también reporta un coeficiente alfa de Cronbach de 0,806 en una muestra de 112 estudiantes.

En consecuencia, el instrumento no se presenta aquí como una medida “nueva” que deba ser revalidada íntegramente en cada investigación, sino como una versión previamente adaptada y con evidencias psicométricas documentadas, utilizada en este trabajo con una finalidad orientativa y metacognitiva. Las puntuaciones obtenidas mediante el CAMEA40 se incorporan como variables descriptivas en la matriz de análisis, sin emplearse para establecer diagnósticos rígidos, etiquetas fijas o explicaciones causales del abandono universitario.

El cuarto eje es la analítica del aprendizaje, entendida como el uso sistemático de datos educativos para describir trayectorias, identificar patrones y apoyar decisiones pedagógicas. En este estudio, el análisis de datos educativos se combina con técnicas de aprendizaje automático para explorar variables asociadas con el abandono universitario. La matriz de datos integra información procedente del cuestionario, de la trayectoria educativa previa, de características personales y familiares y de los registros académicos del primer año. Los modelos predictivos se utilizan aquí con una finalidad exploratoria y de apoyo a la orientación. Las revisiones recientes sobre analítica del aprendizaje apoyada por IA subrayan su potencial para favorecer el desarrollo metacognitivo y socioemocional (Pacheco et al., 2025), siempre que se diseñe con sensibilidad pedagógica y orientación centrada en el estudiante.

La articulación de estos cuatro ejes permite comprender el curso como un dispositivo de orientación y acompañamiento, y no únicamente como una experiencia de formación digital. El autoconocimiento aporta el contenido de la reflexión; la metacognición permite interpretar y regular el aprendizaje; los estilos de aprendizaje ofrecen un lenguaje orientativo para analizar preferencias y estrategias; mientras que la analítica del aprendizaje proporciona

información complementaria para identificar patrones y orientar intervenciones tempranas.

La predicción del abandono mediante modelos analíticos se presenta como una línea especialmente prometedora para el diseño de intervenciones tutoriales ajustadas y oportunas (Ovtšarenko, 2026). Sin embargo, las asociaciones observadas en los datos no deben interpretarse como relaciones causales. En particular, los modelos predictivos pueden contribuir a priorizar apoyos tutoriales, pero cualquier decisión educativa debe considerar la historia, el contexto y la voz del estudiante.

3. METODOLOGÍA

3.1. Diseño del estudio

La investigación adopta un diseño aplicado, longitudinal, no experimental y exploratorio. El estudio integra dos dimensiones relacionadas. La primera corresponde a la experiencia pedagógica desarrollada en el curso e-learning presentado en la plataforma Moodle de la universidad; la segunda consiste en un análisis predictivo de datos educativos destinado a explorar variables asociadas con el abandono universitario y comparar el desempeño de distintos modelos de clasificación. La experiencia formativa constituye el contexto de recogida de datos, mientras que el análisis basado en aprendizaje automático representa la fase empírica y predictiva del estudio.

Los modelos se utilizan con una finalidad exploratoria y de apoyo a la orientación universitaria. Por tanto, sus resultados no se interpretan como un sistema automático para tomar decisiones tutoriales. La información obtenida pretende contribuir a la identificación temprana de posibles situaciones de riesgo y a la planificación de acciones de acompañamiento contextualizadas.

3.2. Contexto pedagógico y participantes

El estudio se realizó en el marco del curso que combina actividades de reflexión metacognitiva, análisis de los estilos de aprendizaje y orientación académica. Para el análisis predictivo se utilizaron los datos recopilados entre los años 2016 y 2021. La matriz analítica final estuvo constituida por 257 estudiantes que disponían de información válida para las variables incluidas en el estudio. La incorporación de los datos correspondientes a la edición 2020/2021 permitió actualizar el conjunto de observaciones y reentrenar los modelos predictivos descritos en trabajos previos (Pulcini et al., 2022; Pulcini, 2025).

La muestra corresponde a un conjunto institucional de casos disponibles y no a una muestra probabilística de la población universitaria italiana. En consecuencia, los resultados se circunscriben al contexto institucional y a las cohortes analizadas y no se generalizan directamente a otras universidades, titulaciones o grupos de estudiantes. La Tabla 1 presenta la distribución de los 257 estudiantes por curso.

Tabla 1. *Cursos académicos de los estudiantes participantes en la muestra*

Curso	Nº estudiantes
Biología de la Nutrición	98
Arquitectura	1
Ciencias Biológicas (curso impartido exclusivamente en inglés)	7
Biociencias y Biotecnología (curso impartido exclusivamente en inglés)	22
Química	8
Química y Tecnología Farmacéutica	27
Diseño Industrial y Ambiental	6
Farmacia	29
Física (curso impartido exclusivamente en inglés)	1
Jurisprudencia	3
Informática	25
Información Científica sobre Medicina y Ciencias del <i>Fitness</i> y Productos Sanitarios	3
Matemáticas y Aplicaciones	5
Medicina Veterinaria	1
Física	3
Ciencias de la Arquitectura	2
Ciencias Gastronómicas	4
Ciencias Sociales para Organizaciones sin Ánimo de Lucro y Cooperación Internacional	4
Seguridad de la Producción Ganadera y Mejora de los Alimentos Típicos de Origen Animal	5
Tecnología y Diagnóstico para la Conservación y Restauración	2
Gestión de los Fenómenos Migratorios y las Políticas de Integración en la Unión Europea	1

3.3. Instrumento y fuentes de información

Una de las fuentes de información fue el, desarrollado a partir del marco teórico del Cuestionario Honey-Alonso de Estilos de Aprendizaje (Chaea) (Alonso et al., 1994). El instrumento permite obtener puntuaciones correspondientes a los estilos activo, reflexivo, teórico y pragmático.

La versión italiana del CAMEA40 fue traducida y adaptada al contexto lingüístico y sociocultural italiano mediante un proceso progresivo de revisión y experimentación con docentes y estudiantes. Este proceso y las aplicaciones iniciales del instrumento en contextos universitarios fueron documentados previamente (Pulcini et al., 2018). En el presente estudio se utiliza la versión italiana previamente adaptada y evaluada, por lo que no se repite una validación psicométrica completa del cuestionario. Las evidencias psicométricas documentadas previamente constituyen el fundamento para su utilización en esta investigación.

El CAMEA40 se emplea como un instrumento de orientación y reflexión metacognitiva, y sus puntuaciones se incorporan como variables descriptivas relacionadas con los estilos de aprendizaje. Además del CAMEA40, se utilizaron datos de la trayectoria educativa previa, información personal y familiar y registros académicos correspondientes al primer año universitario.

3.4. Organización del curso y actividades

Las actividades propuestas en el curso constituyen fuentes de información e incluyen cuestionarios, actividades realizadas en la plataforma Moodle, informes autobiográficos, presentaciones personales y documentos producidos por el estudiantado en tareas de autoanálisis.

El curso se organiza en siete etapas progresivas, concebidas para acompañar al estudiantado en un recorrido formativo que va desde la comprensión conceptual inicial hasta la aplicación práctica y la reflexión final. En la primera fase se abordan los fundamentos vinculados con la neurociencia, la orientación y la metacognición; posteriormente, se introducen los estilos de aprendizaje y se desarrollan de manera diferenciada los perfiles activo, reflexivo, teórico y pragmático. La secuencia culmina con una etapa de síntesis y cierre, orientada a integrar los aprendizajes adquiridos y a promover una revisión crítica del propio proceso de aprendizaje.

A lo largo de los años, las actividades asociadas a cada módulo se han ido ajustando y enriqueciendo en función de la evolución del curso y de la incorporación de nuevos recursos, entre ellos herramientas de IA, que amplían las

posibilidades de elaboración, reflexión y producción académica. En particular, las reflexiones metacognitivas guiadas y el uso del CAMEA40 ayudan al estudiantado a identificar parámetros relevantes sobre su modo de aprender, que posteriormente se emplean, por ejemplo, para elaborar análisis de tipo FODA (Fortalezas, Oportunidades, Debilidades, y Amenazas) o para definir proyectos personales siguiendo el modelo Smart (objetivos específicos, medibles, alcanzables, relevantes y acotados en el tiempo). El alumnado recibe explicaciones y orientaciones para utilizar la IA como apoyo en la construcción de estos productos de autoanálisis.

3.5. Construcción de la matriz de características

Para el análisis predictivo se construyó una matriz de características compuesta por 257 observaciones y 22 variables predictoras, organizadas en cuatro grupos:

- Variables relacionadas con los estilos de aprendizaje: puntuaciones correspondientes a los estilos activo, reflexivo, teórico y pragmático obtenidas mediante el CAMEA40.
- Variables de trayectoria educativa previa: calificación final de la educación secundaria, tipo de institución de procedencia y experiencias educativas o laborales previas relacionadas con la universidad.
- Variables personales y familiares: sexo, edad, nacionalidad, nivel educativo del padre, nivel educativo de la madre y número de hermanos o hermanas.
- Variables de trayectoria académica universitaria: condición de estudiante trabajador, número de exámenes realizados, créditos ECTS obtenidos y promedio de calificaciones durante el primer año.

Antes del entrenamiento de los modelos se revisó la integridad de los registros y se seleccionaron únicamente los casos con información válida para las variables incluidas. Las variables categóricas y numéricas fueron procesadas de acuerdo con los requerimientos de cada algoritmo.

3.6. Definición de la variable objetivo

La variable objetivo fue el estado académico final del estudiante. Se definió como clase positiva el abandono universitario y como clase negativa la finalización de los estudios.

De los 257 estudiantes incluidos en la matriz analítica, 87 fueron clasificados como casos de abandono, lo que representa el 33,9% de la muestra. Los 170 estudiantes restantes finalizaron la trayectoria académica considerada, equivalentes al 66,1%. Ambas categorías representan el 100% de los casos analizados.

La categoría “abandono” se refiere al estado académico registrado en la base de datos y no implica una explicación causal de la interrupción de los estudios. El vector objetivo correspondiente a la variable de interés es el abandono estudiantil.

3.7. Modelos de clasificación

La matriz de características y el vector objetivo se utilizaron para entrenar seis modelos de clasificación supervisada: regresión logística; máquinas de vectores de soporte (SVM); K-vecinos más cercanos (KNN); Random Forest; Naive Bayes; y red neuronal artificial de tipo perceptrón multicapa (MLP).

La selección de estos modelos permitió comparar diferentes familias de algoritmos. La regresión logística representa un enfoque lineal; mientras que SVM permite establecer fronteras de decisión mediante funciones kernel; por su parte, KNN clasifica las observaciones según la proximidad entre casos; Random Forest integra múltiples árboles de decisión; Naive Bayes utiliza un enfoque probabilístico; y el MLP representa un modelo basado en redes neuronales artificiales.

Random Forest y Naive Bayes fueron incorporados en la actualización del análisis, ya que no habían sido utilizados en la versión anterior del estudio. Su desempeño se comparó con el de los modelos considerados previamente. Estos tipos de modelos basados en aprendizaje automático se han utilizado en otros estudios para la predicción temprana de la deserción universitaria (Castro-Maldonado et al., 2026; Aco Tito et al., 2023).

3.8. Reentrenamiento y optimización de hiperparámetros

Los modelos predictivos fueron reentrenados incorporando los datos correspondientes a la edición 2020/2021, con el objetivo de actualizar y fortalecer el análisis realizado en versiones anteriores del proyecto.

La búsqueda de los hiperparámetros óptimos se realizó mediante la función GridSearchCV de la biblioteca Scikit-learn. Esta función permitió evaluar sistemáticamente distintas combinaciones de parámetros para cada algoritmo y seleccionar la configuración con mejor desempeño dentro del procedimiento de validación cruzada (Pedregosa et al., 2011).

En cada modelo se exploraron los hiperparámetros específicos de su técnica. En Random Forest se ajustaron el número de árboles, la profundidad máxima y el tamaño mínimo de las muestras; en SVM, el kernel, el parámetro de regularización, el grado y el valor de gamma; en KNN, la métrica de distancia, el número de vecinos y el sistema de ponderación; en Naive Bayes, el parámetro de suavizado de la varianza; y, finalmente, en el MLP, la arquitectura de la red, el optimizador y los parámetros relacionados con la tasa de aprendizaje.

3.9. Procedimiento de validación

Para verificar el desempeño de cada técnica se implementó la estrategia de *repeated k-fold cross-validation*. El conjunto de datos se dividió en diez particiones (*folds*); donde, en cada iteración, nueve *folds* se utilizaron para entrenar el modelo y un *fold* se reservó como conjunto de prueba para evaluar la capacidad predictiva respecto de la variable objetivo. Este procedimiento se repitió diez veces, cambiando en cada ocasión el *fold* utilizado como prueba, con el fin de obtener estimaciones más robustas y estables del rendimiento del modelo.

La validación cruzada se apoyó en GridSearchCV, de modo que la selección de hiperparámetros y la estimación del rendimiento se realizaron de forma sistemática y repetida sobre distintas particiones del conjunto de datos. Los resultados finales de cada técnica se calcularon a partir del promedio de las prestaciones obtenidas en las iteraciones realizadas.

3.10. Evaluación del desempeño

El desempeño de los seis modelos se evaluó mediante cuatro medidas derivadas de la matriz de confusión: exactitud (*accuracy*), precisión (*precision*), sensibilidad (*recall*) y puntuación F1. La clase positiva correspondió al abandono universitario, por lo que la sensibilidad se consideró una métrica especialmente relevante para valorar la capacidad del modelo para identificar los casos de abandono, aunque la interpretación final se realizó considerando conjuntamente las cuatro medidas.

3.11. Tratamiento ético y protección de datos

Los datos utilizados fueron tratados con fines académicos y de investigación. Para la construcción de la matriz de características se utilizaron registros codificados y no se incorporaron nombres ni identificadores personales directamente reconocibles.

El acceso a la información se restringió al equipo responsable del estudio y los resultados se presentan de forma agregada. El análisis se desarrolló atendiendo a los principios de confidencialidad, minimización de datos y uso responsable de la información académica. Sus resultados se consideran información complementaria para orientar futuras acciones de tutoría y acompañamiento, siempre bajo supervisión humana y mediante una interpretación contextualizada.

3.12. Consideraciones metodológicas

Debe mencionarse que el tamaño y la composición de la muestra limitan la generalización de los resultados. Los 257 casos proceden de una experiencia institucional concreta y fueron seleccionados a partir de la disponibilidad de información válida. Además, la distribución entre abandono y finalización no es equivalente, por lo que se analizaron conjuntamente medidas como precisión, sensibilidad y F1-score, evitando depender exclusivamente de la exactitud.

Además, la ausencia de una cohorte externa independiente limita la comprobación de la estabilidad de los resultados en otros contextos. Por ello, los valores de desempeño deben interpretarse como resultados preliminares. Futuras investigaciones deberán ampliar la muestra, incorporar nuevas cohortes y evaluar la estabilidad de los modelos mediante validación externa.

La utilización del CAMEA40 se fundamenta en la adaptación y las evidencias psicométricas documentadas previamente (Pulcini et al., 2018). No se realiza aquí una validación psicométrica ya que el objetivo principal es comparar modelos de clasificación y explorar asociaciones con el abandono universitario.

4. RESULTADOS

Los resultados obtenidos permiten distinguir dos planos complementarios: por un lado, los efectos formativos asociados al curso *e-learning* y, por otro, los hallazgos derivados del análisis predictivo. En el plano formativo, el cuestionario final del curso muestra una valoración global positiva por parte del estudiantado, que reconoce la utilidad del trabajo realizado para comprender mejor sus propias formas de aprender, reflexionar sobre su trayectoria académica y vincular los contenidos del curso con su desarrollo personal y profesional. Las actividades orientadas al autoconocimiento y a la reflexión metacognitiva fueron especialmente valoradas por favorecer una lectura más consciente

del propio proceso de aprendizaje. Asimismo, se identificó un interés por los módulos y actividades que favorecen la conexión entre teoría y práctica.

En particular, las reflexiones metacognitivas guiadas y el uso del cuestionario CAMEA40 ayudan al estudiantado a identificar parámetros relevantes sobre su modo de aprender, que posteriormente se emplean para elaborar productos de autoanálisis. El alumnado valora especialmente las actividades de autoanálisis apoyadas por IA, que incluyen la elaboración de productos de reflexión y la formulación de objetivos personales que transforman metas vagas en planes concretos y verificables.

Estas experiencias parecen favorecer una articulación más clara entre autoconocimiento y acción. La secuencia exploración–reflexión–comprensión–decisión permite que el estudiantado no se limite a recibir resultados, sino que los interprete y los convierta en un plan de trabajo personal. Este dato sugiere que la IA resulta más valiosa cuando se integra en una arquitectura pedagógica que orienta la reflexión y evita un uso superficial o meramente instrumental. En conjunto, los resultados del cuestionario final de curso indican que la experiencia formativa ha cumplido adecuadamente sus objetivos formativos.

En el plano analítico, se realizaron análisis exploratorios mediante el cálculo del coeficiente de correlación de Pearson entre las 22 variables incluidas en la matriz de características y el vector objetivo correspondiente al abandono universitario. La muestra analizada estuvo compuesta por 257 casos válidos: 87 estudiantes en situación de abandono y 170 estudiantes que finalizaron la trayectoria académica considerada. En la Tabla 2 se presentan las correlaciones de Pearson obtenidas para cada variable en relación con el abandono.

El coeficiente de Pearson adopta valores entre -1 y $+1$: un valor cercano a cero indica ausencia de correlación lineal, mientras que valores próximos a $+1$ o -1 señalan una correlación positiva o negativa, respectivamente. La magnitud del coeficiente permite estimar la intensidad de la relación entre las variables

Para la interpretación de los resultados, se adoptó la clasificación propuesta por Dancey y Reidy (2019), que distingue entre correlaciones nulas, débiles, moderadas, fuertes y perfectas, para el ámbito de la investigación psicológica. Asimismo, las variables se organizaron en la tabla según el valor del coeficiente de Pearson, de mayor correlación positiva a mayor correlación negativa.

Tabla 2. *Matriz de correlación de características en referencia al vector objetivo de abandono*

Característica	Coefficiente de correlación (Pearson)	Clasificación*
Edad	0.24	Débil
Escuela secundaria italiana profesional que se aproxima a la Formación profesional de grado medio española	0.14	Débil
Escuela secundaria italiana que se aproxima al Bachillerato español de Humanidades y Ciencias sociales	0.13	Débil
Escuela secundaria italiana que se aproxima al Bachillerato español de Artes	0.12	Débil
Número de hermanos y/o hermanas	0.10	Débil
Estilo activo	0.10	Débil
Experiencias educativas previas en la universidad	0.07	Cero
Estudiante/trabajador	0.06	Cero
Nivel de estudio de la madre	0.05	Cero
Sexo	0.04	Cero
Experiencia laboral previa en la universidad	0.04	Cero
Nivel de estudio de los padres	0.03	Cero
Estilo pragmático	-0.01	Cero
Estilo reflexivo	-0.03	Cero
Escuela secundaria italiana técnica que se aproxima a la Formación profesional de grado superior española	-0.03	Cero
Estilo teórico	-0.10	Débil
Nacionalidad italiana	-0.17	Débil
Escuela secundaria italiana que se aproxima al Bachillerato español de Ciencias	-0.19	Débil
Calificación final escuela secundaria italiana	-0.25	Débil
Promedio de calificaciones del primer año de universidad	-0.48	Moderado
Total de exámenes cumplidos en el primer año de universidad	-0.66	Fuerte
Total de créditos ECTS (Sistema Europeo de Transferencia y Acumulación de Créditos) cursados durante el primer año de universidad	-0.73	Fuerte

Desde el punto de vista predictivo, el análisis realizado con una muestra de 257 estudiantes permitió identificar un conjunto de variables asociadas con una mayor probabilidad de abandono. Entre los factores con asociación positiva con la deserción se observaron la edad, la procedencia de escuelas de formación profesional, el itinerario de Bachillerato en Humanidades y Ciencias sociales, el Bachillerato de Artes, el número de hermanos y la preferencia por un estilo de aprendizaje activo. Por el contrario, mostraron una asociación negativa con el abandono el estilo teórico, la nacionalidad italiana, la asistencia a bachilleratos de Ciencias, la calificación final de la escuela secundaria, el promedio de calificaciones del primer año universitario, el número de exámenes rendidos durante dicho período y los créditos obtenidos en el primer año.

Estos resultados merecen una lectura prudente. No establecen relaciones causales, pero sí sugieren que el riesgo de abandono no puede explicarse a partir de una sola dimensión. La combinación de variables académicas, socio-demográficas y disposicionales refuerza la idea de que la permanencia universitaria constituye un fenómeno multifactorial, en el que se entrecruzan antecedentes escolares, condiciones de vida, hábitos de estudio y trayectorias de integración institucional (Bernardo et al., 2022).

A continuación, en la Tabla 3 se presentan los hiperparámetros seleccionados para cada uno de los seis modelos de clasificación utilizados en el estudio: regresión logística, SVM, K-vecinos más cercanos, Random Forest, Naïve Bayes y perceptrón multicapa (MLP). Los parámetros finales fueron determinados mediante GridSearchCV, dentro del procedimiento de validación cruzada repetida.

Tabla 3. *Hiperparámetros establecidos para cada técnica utilizada para el entrenamiento del modelo predictivo relacionado con el abandono*

Técnica del Modelo	Hiperparámetros
Regresión logística	solver = 'liblinear', penalty='l1', C = 1 max_iter= 600
SVM	kernel='poly', gamma= 1000,C = 0,001 grado = 1
K-Más cercano	metric= 'cosine', n_neighbors = 20, weights='distance'
Random Forest	n_estimators=670, min_samples_split=2, min_samples_leaf=2, max_features='auto', max_depth=70, bootstrap=True

Técnica del Modelo	Hiperparámetros
Naive Bayes	priors=None var_smoothing=0.0533669923120631
Red neuronal MLP	capa de entrada = 22 nodos, capa oculta 1 = 60 nodos capa oculta 2 = 30 nodos capa oculta 3 = 20 nodos capa de salida = 1 nodo optimizador = Adam, learning_rate=0.04,beta_1=0.999, beta_2=0.999, epsilon=1e-05, amsgrad=False.

El desempeño de los modelos predictivos, obtenidos con las seis técnicas anteriormente mencionadas, se evaluó con cuatro medidas de calidad obtenidas de la matriz de confusión. En dicha matriz se denominan como positivos a los estudiantes que abandonan y como negativos a los que logran el título de graduación, a su vez se identifican verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN). Se calculan las siguientes medidas de calidad (Vinces-Vinces & Flores-Sanchez, 2025):

- Exactitud, que es el número de predicciones correctas / número total de muestras en el conjunto de datos, es decir: $(TP+TN) / (TP + TN+ FP+FN)$.
- Precisión, que permite estimar la capacidad del clasificador para no etiquetar como positiva una muestra que es negativa (no etiquetar como “abandono” a un estudiante que ha terminado el curso); es decir: $TP / (TP + FP)$.
- Sensibilidad: $TP / (TP + FN)$, la sensibilidad por lo tanto permite medir la capacidad del clasificador para identificar todos los casos positivos presentes (todos los casos de abandono);
- Puntuación, F1 (media armónica entre precisión y sensibilidad), calculada como $F1 = 2(precisión \times sensibilidad)/(precisión + sensibilidad)$.
- Los indicadores de desempeño se han calculado tomando como referencia la clase objetivo correspondiente al abandono estudiantil. En consecuencia, para la interpretación de los resultados se han considerado la exactitud, la precisión, la sensibilidad y el F1-score asociados a dicha clase. A continuación, se muestra la Tabla 4 con los resultados obtenidos por cada técnica.

Tabla 4. *Resultados obtenidos por cada técnica utilizada para el entrenamiento*

Técnicas/Medidas de calidad	Regresión logística	SVM	K-Más cercano	Red neuronal MLP	Random Forest	Naive Bayes
Exactitud	0,88	0,88	0,83	0,87	0,91	0,84
Precisión	0,92	0,93	0,90	0,83	0,91	0,84
Sensibilidad	0,75	0,72	0,61	0,80	0,83	0,68
Puntuación F1	0,82	0,80	0,72	0,81	0,86	0,74

Los resultados indican un rendimiento general elevado en todos los modelos, aunque con diferencias apreciables entre ellos. Así, Random Forest obtuvo el mejor desempeño global, con una exactitud de 0,91, una precisión de 0,91, una sensibilidad de 0,83 y una puntuación F1 de 0,86. La red neuronal MLP y la regresión logística también alcanzaron resultados satisfactorios, mientras que SVM, K-vecinos más cercanos y Naive Bayes mostraron valores algo más moderados, especialmente en sensibilidad.

De forma general, los resultados sugieren que los modelos basados en estructuras no lineales y de ensamblado presentan una mayor capacidad para identificar los casos de abandono en esta muestra. Sin embargo, debido al carácter exploratorio del estudio y a la ausencia de una cohorte externa independiente, estos resultados deben interpretarse como preliminares.

Estos resultados vinculan el análisis predictivo con el objetivo formativo central del curso que consistía en identificar, entre el estudiantado de primero y segundo año, los perfiles asociados a un mayor riesgo que puedan beneficiarse de una tutoría específica en los primeros momentos de su trayectoria formativa.

5. CONCLUSIONES

La integración de los datos presentados permite sostener que el curso “Conócete a ti mismo. Descubre tu estilo de aprendizaje” opera en dos niveles complementarios. Por un lado, promueve los procesos de autoconocimiento, metacognición y toma de decisiones mediante actividades estructuradas y mediadas tecnológicamente. Por otro, utiliza la inteligencia artificial y la analítica del aprendizaje para explorar patrones de riesgo y orientar intervenciones tempranas. Esta doble dimensión fortalece la coherencia del proyecto, ya que articula prevención pedagógica y análisis predictivo en un mismo dispositivo formativo.

Los resultados también sugieren que la IA adquiere valor educativo cuando se integra en una lógica tutorial y no solo técnica. El hecho de que

determinadas actividades asistidas por IA se orienten al análisis del propio perfil de aprendizaje, a la elaboración de productos de reflexión (incluidos los análisis de tipo FODA, la formulación de objetivos Smart y la interpretación crítica de los datos), muestra que la tecnología funciona como mediación para el aprendizaje autorregulado más que como simple automatización. En este sentido, el curso evita una lectura determinista de los datos y propone un uso crítico de la predicción, coherente con la idea de que predecir no significa etiquetar, sino abrir oportunidades de transformación.

La experiencia descrita no se limita a una secuencia de contenidos; más bien propone un acompañamiento progresivo que parte de la exploración personal y culmina en la elaboración de metas concretas. Esta lógica resulta especialmente valiosa en los primeros años universitarios, cuando el estudiantado todavía está construyendo hábitos de estudio, expectativas realistas y una identidad académica más estable, en consonancia con la literatura reciente sobre pertenencia, permanencia y compromiso en educación superior. En ese sentido, el curso funciona como una mediación para pensar el propio recorrido, reconocer dificultades sin dramatizarlas y activar recursos que favorecen una participación más autónoma en la vida universitaria (Allen et al., 2024).

La incorporación de inteligencia artificial, lejos de reemplazar la relación educativa, amplía las posibilidades de análisis, visualización y síntesis. Bien orientada, puede apoyar la elaboración de productos académicos, facilitar la personalización de tareas y contribuir a una comprensión más fina de patrones de comportamiento y aprendizaje, como sugieren revisiones recientes sobre inteligencia artificial y analítica del aprendizaje centradas en el estudiante. Sin embargo, este potencial solo se traduce en valor educativo cuando está guiado por criterios de sentido pedagógico, inclusión y acompañamiento humano, ya que la evidencia destaca la necesidad de enfoques centrados en las personas, con participación docente y de los distintos actores educativos, así como con criterios pedagógicos claros (Alfredo et al., 2024).

Asimismo, el análisis predictivo aporta un criterio de alerta temprana que permite actuar antes de que la desvinculación se consolide. El interés de estos modelos no reside únicamente en clasificar correctamente, sino en ofrecer a las instituciones una base para intervenir con mayor pertinencia y oportunidad, tal y como lo demuestran los estudios recientes sobre detección temprana y modelos de riesgo de deserción. Por ello, estos enfoques analíticos deben entenderse como una herramienta de apoyo a la tutoría, no como un dispositivo autónomo de decisión, especialmente cuando la investigación insiste en marcos de uso responsables, con supervisión humana y orientación institucional (Hoyos & Daza, 2023).

Desde una perspectiva comparada, los resultados obtenidos para identificar características asociadas con una mayor probabilidad de abandono se alinean con la literatura previa sobre deserción universitaria. La evidencia disponible ha mostrado que los factores asociados al ingreso a la universidad, el rendimiento académico previo y determinadas variables sociodemográficas influyen de manera significativa en la permanencia estudiantil, mientras que el desempeño académico acumulado constituye uno de los predictores más consistentes del abandono (Barrios, 2011; Constante-Amores et al., 2020). Asimismo, investigaciones recientes han mostrado que variables vinculadas con la elección de los estudios y con la adecuación entre el perfil del estudiante y la trayectoria formativa también inciden de manera significativa en la permanencia universitaria (Constante-Amores et al., 2020). En consecuencia, los hallazgos del presente estudio refuerzan la pertinencia de considerar de manera integrada variables personales, académicas e institucionales en el análisis del riesgo de deserción.

Desde la perspectiva de la permanencia universitaria, la identificación de variables asociadas al abandono y el buen desempeño del modelo Random Forest ofrecen una base empírica valiosa para orientar futuras acciones institucionales. Los métodos basados en árboles parecen captar mejor las relaciones entre las variables predictoras y la variable objetivo, lo que resulta coherente con los contextos caracterizados por interacciones no lineales y las relaciones complejas entre factores personales, académicos y de trayectoria formativa. No obstante, estos resultados deben interpretarse con prudencia, ya que las correlaciones observadas no agotan la complejidad de las trayectorias estudiantiles y requieren una lectura pedagógica contextualizada.

En esta línea, las futuras investigaciones deberían ampliar el conjunto de datos para reforzar la validación de las técnicas de aprendizaje automático e incorporar nuevos algoritmos predictivos, como árboles de decisión u otros modelos basados en *ensembles*, con el fin de contrastar su rendimiento con el de los modelos ya evaluados. Del mismo modo, convendría enriquecer la matriz de características mediante la integración de otros datos estructurados del curso *e-learning* (por ejemplo, con cuestionarios sobre ansiedad, resiliencia, enfoque de estudio y variables afines), lo que permitiría comparar su peso correlacional y predictivo con los hallazgos de este estudio preliminar.

En conjunto, estas líneas de trabajo podrían fortalecer el uso pedagógico de la inteligencia artificial como una herramienta de apoyo para experiencias más inclusivas, orientadoras y transformadoras, con el objetivo de no dejar a nadie atrás.

AGRADECIMIENTO

El autor expresa su más sincero agradecimiento a la Universidad de Camerino (Italia) por el apoyo financiero y los recursos proporcionados para la realización de esta investigación. Asimismo, expresa su especial agradecimiento a L. Cappellacci, S. Boria, L. Cesaretti y V. Polzonetti por su colaboración y apoyo.

REFERENCIAS BIBLIOGRÁFICAS

- Aco Tito, E. A., Hanco Condori, B. O., & Pérez Vera, Y. (2023). Análisis comparativo de técnicas de Machine Learning para la predicción de casos de deserción universitaria. *Risti, Revista Ibérica de Sistemas e Tecnologías de Informação*, (51), 84-98. <https://doi.org/10.17013/risti.51.84-98>
- Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 6. <https://doi.org/10.1016/j.caeai.2024.100215>
- Allen, K.-A., Slaten, C., Hong, S., Ma, L., Craig, H., May, F., & Counted, V. (2024). Belonging in higher education: A twenty-year systematic review. *Journal of University Teaching and Learning Practice*, 21(5), 1-55. <https://doi.org/10.53761/s2he6n66>
- Alonso, C., Gallego, D., & Honey, P. (1994). *Los estilos de aprendizaje. Procedimientos de diagnóstico y mejora*. Mensajero.
- Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca. (26 de marzo, 2026). ANVUR publishes the 2026 report on the higher education and research system. *Home, Press releases*. <https://www.anvur.it/en/press-releases/anvur-publishes-2026-report-higher-education-and-research-system>
- Barrios, A. B. (2011). Deserción universitaria en Chile: Incidencia del financiamiento y otros factores asociados. *Revista CIS*, 9(14), 59-72.
- Bernardo, A. B., Galve-González, C., Núñez, J. C., & Almeida, L. S. (2022). A path model of university dropout predictors: The role of satisfaction, the use of self-regulation learning strategies and students' engagement. *Sustainability*, 14(3). <https://doi.org/10.3390/su14031057>
- Castro-Maldonado, J. J., Londoño-Gallego, J. A., Rodríguez-Marín, P. A., & Martínez-Vargas, J. D. (2026). Applicazione di modelli di Machine Learning per la previsione dell'abbandono scolastico degli studenti di un istituto colombiano di formazione per competenze. *Comunicare*, 34(85), 171-202. <https://doi.org/10.5281/zenodo.19690782>

- Constante-Amores, A., Florenciano Martínez, E., Navarro Asencio, E., & Fernández-Mellizo, M. (2020). Factores asociados al abandono universitario. *Educación XXI*, 24(1). <https://doi.org/10.5944/educxx1.26889>
- Dancey, C. P., & Reidy, J. (2019). *Estatística sem matemática para psicologia* (7.ª ed.). Penso.
- Hoyos, J. K., & Daza, G. (2023). Modelo predictivo para identificar estudiantes universitarios con alto grado de deserción. *Revista Electrónica de Investigación Educativa*, 25. <https://doi.org/10.24320/redie.2023.25.e13.5398>
- Madrigal Gil, A. J., & Trujillo Torres, J. M. (2017). CAMEA40: Una opción para el acompañamiento tutorial. *Revista de Estilos de Aprendizaje*, 10(20). <https://doi.org/10.55777/rea.v10i20.1063>
- Ovtšarenko, O. (2026). Using AI to forecast student dropout risk in technical education using a learning analytics approach. *Scientific Reports*, 16. <https://doi.org/10.1038/s41598-026-44919-1>
- Pacheco, A. J., Boude Figueredo, O. R., Chiappe, A., & Fontán de Bedout, L. (2025). AI-powered learning analytics for metacognitive and socio-emotional development: A systematic review. *Frontiers in Education*, 10. <https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2025.1672901>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., & Louppe, G. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. <https://jmlr.org/papers/v12/pedregosa11a.html>
- Pulcini, G. G. (2025). Conócete a ti mismo: Herramientas metacognitivas para comprender y optimizar tus estilos de aprendizaje. En M. V. Afonso, M. Giuliano, S. Pérez, & R. B. Vallejos (Comps.), *Enseñanza e innovación* (pp. 96-111). EDUNO.
- Pulcini, G. G., Polzonetti, V., Cappellacci, L., & Cesaretti, L. (2022). Prevenire l'abbandono universitario: Stili di apprendimento e machine learning. En J. M. Esteve Faubel, A. Fernández-Sogorb, R. Martínez-Roig, & J.-F. Álvarez-Herrero (Eds.), *Transformando la educación a través del conocimiento* (pp. 1046-1060). Ediciones Octaedro. <https://octaedro.com/wp-content/uploads/2022/11/9788419506733.pdf>
- Pulcini, G. G., Porcarelli, A., Angeletti, M., & Polzonetti, V. (2018). Alla ricerca di strumenti per l'orientamento universitario: Prima traduzione ufficiale del Cuestionario adaptado de estilos de aprendizaje (CAMEA40) in lingua italiana. *Sperimentazione avviata all'Università di Camerino. IUSVEducation*, (11), 174-205.

- <https://iusveducation.it/alla-ricerca-di-strumenti-per-lorientamento-universitario-prima-traduzione-ufficiale-del-cuestionario-adaptado-de-estilos-de-aprendizaje/>
- Vinces-Vinces, F. V., & Flores-Sánchez, M. (2025). Aplicación de Machine Learning para predecir y explicar el rendimiento académico universitario. *Comunicar*, 33(82), 12-23. <https://doi.org/10.5281/zenodo.16121350>
- Zapata-Zapata, A., Vesga Bravo, G. J., Puente Ferreras, A., & Alvarado Izquierdo, J. M. (2024). Juicios metacognitivos en los procesos de aprendizaje en la educación superior: una revisión sistemática 2018-2023. *Revista Latinoamericana de Psicología*, 56(15), 147-155. <https://doi.org/10.14349/rlp.2024.v56.15>

Cómo citar este artículo: Pulcini, G. (2026). Estilos de aprendizaje y abandono universitario: un enfoque metacognitivo y de analítica predictiva. *Educación*, XXXV(69), 32-53. <https://doi.org/10.18800/educacion.202602.M003>

Primera publicación: 14 de septiembre de 2026.

Este es un artículo de acceso abierto distribuido bajo los términos de Licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0), que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.