

Análisis de información generada en las plataformas educativas a partir de herramientas de Inteligencia Artificial



**Universidad
Internacional
de Valencia**

00 Mes 2026

Titulación:

MIAR – Máster universitario
en inteligencia artificial

Curso académico

2025 – 2026

Alumno/a:

Michael Laudrup Luis González

Director/a de TFM: Irma

Zoraida Sanabria Cárdenas

De:

 Planeta Formación y Universidades

Índice

Resumen	6
Abstract.....	8
1. Introducción	10
2. Objetivos	12
2.1. Objetivo General	12
2.2. Objetivos Específicos	12
3. Estado del Arte y Marco teórico.....	15
3.1. Marco teórico.....	15
3.1.1. Minería de Datos Educativos (EDM) y Learning Analytics.....	15
3.1.2. Heterogeneidad de datos y estándares	16
3.1.3. Deserción frente a bajo rendimiento	18
3.1.4. Aprendizaje No Supervisado: Descubrimiento de Patrones (Clustering) 18	
3.1.5. Aprendizaje Supervisado y Modelos de Ensamble	19
3.1.6. Conclusiones	20
3.2. Estado del arte.	20
3.2.1. Aprendizaje no supervisado en EDM.....	21
3.2.2. Aprendizaje supervisado en EDM.	22
3.2.3. El enfoque híbrido.....	23
3.2.4. Inteligencia Artificial Explicable (XAI) en Educación.	24
3.2.5. Conclusiones	25
4. Desarrollo del proyecto y resultados	27
4.1. Metodología.	27
4.1.1. Metodología CRISP-MD.	27
4.1.2. Definición de tareas.....	29

4.1.3.	Planificación.	31
4.2.	Planteamiento del problema.	¡Error! Marcador no definido.
4.3.	Desarrollo del proyecto	¡Error! Marcador no definido.
4.4.	Resultados	¡Error! Marcador no definido.
5.	Conclusión y trabajos futuros	¡Error! Marcador no definido.
6.	Referencias	32
	Apéndice I.....	34
	Anexos I.....	35

iii **INFORMATIVO!!!** (quitar)

**Actualizar tabla de contenidos*

**Hacer uso de salto de página para los nuevos apartados.*

**Hacer uso de los estilos definidos: Normal, Título1, Título2, Título3.*

Índice de ilustraciones

Ilustración 1 Proceso de preparación del dataset OULAD. (Kuzilek, Hlosta, & Zdrahal, 2017)..... 17

Ilustración 2. Arquitectura Jupyter Cliente - Servidor. Fuente: <https://www.paradigmadigital.com/dev/jupyter-data-science-aplicada/>¡Error!
Marcador no definido.

Ilustración 3. Cronograma de las tareas definidas. Elaboración propia.¡Error!
Marcador no definido.

¡¡¡INFORMATIVO!!! (quitar)

**Actualizar índice de tablas (Referencias / Insertar tabla de ilustraciones)*



Índice de tablas

¡Error! Marcador no definido.

iii **INFORMATIVO!!!** (quitar)

**Actualizar índice de tablas (Referencias / Insertar tabla de ilustraciones)*

Resumen

En la actualidad, los Sistemas de Gestión del Aprendizaje (LMS) se han consolidado como la infraestructura fundamental del entorno educativo digital. Si bien estas plataformas han facilitado la generación de un volumen masivo de datos, persiste una brecha crítica entre la recolección de estos y su aprovechamiento pedagógico. En la práctica, los LMS operan frecuentemente como "cajas negras", almacenando registros que los docentes no pueden interpretar, lo que impide la detección temprana de estudiantes en riesgo de fracaso.

El presente Trabajo de Fin de Máster (TFM) aborda esta problemática mediante el desarrollo y validación de un modelo híbrido de Minería de Datos Educativos (EDM). La investigación se fundamenta en la explotación del conjunto de datos OULAD (*Open University Learning Analytics Dataset*), uno de los repositorios de referencia más completos a nivel global, que integra información demográfica, resultados de evaluaciones y registros detallados de interacción de miles de estudiantes en diversos cursos universitarios.

El núcleo del estudio se centra en la aplicación práctica de técnicas de Inteligencia Artificial sobre estos datos. Tras una fase de Ingeniería de Características (*Feature Engineering*) para transformar los *logs* brutos en métricas de comportamiento pedagógico, se implementan estrategias de aprendizaje no supervisado mediante algoritmos de agrupamiento se segmenta a la población estudiantil para descubrir arquetipos de aprendizaje y patrones de conducta sin etiquetas previas.

Posteriormente, estos hallazgos alimentan modelos de aprendizaje supervisado. El objetivo es entrenar sistemas predictivos capaces de anticipar el rendimiento final del alumno. Para garantizar que estas predicciones sean útiles en un entorno real, se integran técnicas de Inteligencia Artificial Explicable (XAI), desvelando qué variables determinan el riesgo académico. De este modo, el sistema no solo predice el fracaso, sino que ofrece al docente las claves para realizar intervenciones personalizadas y fundamentadas.

Palabras clave: Minería de Datos Educativos, Moodle, Aprendizaje Supervisado, Aprendizaje no supervisado, Analítica de Aprendizaje, OULAD, Inteligencia Artificial Explicable.

Abstract

Currently, Learning Management Systems (LMS) have established themselves as the fundamental infrastructure of the digital educational environment. Although these platforms have facilitated the generation of massive volumes of data, a critical gap persists between data collection and its pedagogical utilization. In practice, LMS often operate as "black boxes," storing logs that instructors cannot easily interpret, which hinders the early detection of students at risk of academic failure.

This Master's Thesis addresses this problem through the development and validation of a hybrid Educational Data Mining (EDM) model. The research is grounded in the exploitation of the OULAD (Open University Learning Analytics Dataset), a global reference repository that integrates demographic information, assessment results, and detailed interaction logs from thousands of students.

The core of this study focuses on the practical application of Artificial Intelligence techniques to these data. Following a Feature Engineering phase to transform raw logs into pedagogical behavior metrics, unsupervised learning strategies are implemented through clustering algorithms to segment the student population and discover learning archetypes.

Subsequently, these findings feed supervised learning models, including a Transformer-based architecture optimized for capturing long-term sequential dependencies. To ensure these predictions are useful in a real-world setting, Explainable Artificial Intelligence (XAI) techniques are integrated to reveal the variables determining academic risk. Consequently, the system not only predicts failure but also provides instructors with the necessary insights to perform personalized and evidence-based pedagogical interventions.

Keywords: Educational Data Mining, Moodle, Supervised Learning, Unsupervised Learning, Learning Analytics, OULAD, Explainable Artificial Intelligence.

1.Introducción

La educación contemporánea se encuentra inmersa en una transformación impulsada por la digitalización de los procesos de enseñanza y aprendizaje. En el centro de este ecosistema digital se sitúan los Sistemas de Gestión del Aprendizaje (LMS, por sus siglas en inglés), plataformas que han evolucionado de ser repositorios estáticos de documentos hasta convertirse en entornos dinámicos y complejos donde se orquesta la experiencia educativa. Entre estos sistemas, Moodle (Modular Object-Oriented Dynamic Learning Environment) se ha consolidado como el estándar de facto en el ámbito académico global, gracias a su arquitectura de código abierto y su flexibilidad pedagógica, siendo adoptado por instituciones para gestionar cursos que van desde la presencialidad hasta la educación a distancia.

Esta adopción ha traído una generación masiva y continua de datos, conocida como Big Data Educativo. Cada interacción que un estudiante realiza dentro de la plataforma —desde el acceso a un material de lectura, la participación en un foro de discusión, la descarga de una tarea, e incluso la navegación entre menús o la realización de un cuestionario— deja una huella digital registrada en los archivos de registro (*logs*) del sistema.

Sin embargo, a pesar de esta gran abundancia de datos, la infraestructura tecnológica ha avanzado más rápido que la capacidad pedagógica para explotarla. Existe una brecha significativa entre lo que la tecnología registra y lo que el docente conoce realmente sobre el comportamiento de sus estudiantes. Esta situación genera un escenario donde el LMS se convierte en una "caja negra" de información no procesada. Los docentes carecen de herramientas para identificar a tiempo qué alumnos están en riesgo de fracaso o qué perfiles de aprendizaje coexisten en su aula, y a menudo detectan las dificultades cuando ya es demasiado tarde para intervenir. Existe, por tanto, una necesidad crítica de desarrollar sistemas inteligentes que actúen como soporte, procesando masivamente estos registros para revelar

patrones ocultos y ofreciendo información que permita tomar decisiones pedagógicas informadas y oportunas.

El presente Trabajo de Fin de Máster (TFM) se propone abordar esta problemática mediante la aplicación de técnicas avanzadas de Inteligencia Artificial (IA) y Minería de Datos Educativos (EDM). El propósito principal es transformar los datos brutos generados en plataformas de aprendizaje (LMS) en conocimiento accionable, explorando cómo la modelización computacional puede identificar perfiles de comportamiento y predecir el riesgo académico. De este modo, se busca dotar al docente de herramientas analíticas que superen la mera gestión administrativa y faciliten una toma de decisiones pedagógicas fundamentada y proactiva.

Para dar respuesta a esta problemática y cumplir con el propósito de la investigación, la presente memoria se ha organizado en cinco capítulos que estructuran el desarrollo del trabajo. A continuación, se describe el contenido de cada uno de ellos:

Tras esta introducción, el Capítulo 2 define los Objetivos del proyecto, estableciendo tanto la meta general como los hitos específicos que guían la investigación. Posteriormente, el Capítulo 3 aborda el Estado del Arte y Marco Teórico, proporcionando una revisión exhaustiva de la literatura actual y fundamentando los conceptos clave sobre Minería de Datos Educativos y Analítica de Aprendizaje que sustentan la propuesta técnica.

El núcleo del trabajo se desarrolla en el Capítulo 4, titulado Desarrollo del proyecto y resultados. Esta sección desglosa la metodología empleada para abordar el estudio y formaliza el planteamiento del problema específico a resolver. A su vez, detalla las fases de desarrollo del proyecto, desde la extracción de los datos hasta la implementación de los algoritmos, culminando con la exposición y análisis de los resultados obtenidos tras la experimentación.

Finalmente, el Capítulo 5 presenta las Conclusiones y trabajos futuros, donde se sintetizan los hallazgos principales, se discute el impacto de la solución en la labor docente y se proponen nuevas líneas de investigación para dar continuidad al estudio.

El documento cierra con el apartado de Referencias, que recoge las fuentes bibliográficas y recursos académicos citados a lo largo de la memoria.

2. Objetivos

Tras haber analizado la brecha existente entre la recolección masiva de datos en los entornos virtuales y su limitado aprovechamiento pedagógico, es imperativo formalizar las metas que rigen esta investigación. El propósito de este capítulo es delimitar el alcance del proyecto, estableciendo una hoja de ruta que transite desde la fundamentación teórica hasta la implementación de una arquitectura de *Deep Learning* capaz de transformar registros brutos en conocimiento accionable.

Para garantizar el éxito del sistema predictivo, se han definido metas que equilibran la potencia computacional de los modelos híbridos con la transparencia necesaria para el entorno docente. A continuación, se detalla el fin último del estudio y los hitos operativos que permitirán validar la eficacia y explicabilidad del modelo propuesto.

2.1. Objetivo General

Diseñar, implementar y evaluar una arquitectura híbrida de Deep Learning que integre modelos de Autoencoders para la representación de espacios latentes y redes Transformer para el modelado secuencial, con el fin de predecir el riesgo de fracaso y abandono académico en el dataset OULAD y garantizar la interpretabilidad de las decisiones mediante técnicas de Inteligencia Artificial Explicable (XAI).

2.2. Objetivos Específicos

- Analizar la evolución del Estado del Arte en Minería de Datos Educativos, contrastando las limitaciones de los métodos lineales clásicos (PCA, Regresión) frente a las capacidades de las arquitecturas profundas (Deep Learning) para modelar la complejidad del comportamiento estudiantil.

- Consolidar y preprocesar el conjunto de datos OULAD, transformando los registros brutos de interacción (logs) en secuencias temporales estructuradas que permitan capturar la evolución dinámica del aprendizaje, más allá de las métricas estáticas acumuladas. Esta relación temporal es muy importante para las redes neuronales de tipo transformer.
- Implementar una arquitectura de Autoencoders (Aprendizaje de Representaciones) para transformar los datos de alta dimensionalidad en un espacio latente comprimido, capaz de capturar relaciones no lineales que escapan a técnicas tradicionales como el PCA.
- Aplicar algoritmos de Clustering sobre el espacio latente (Aprendizaje No Supervisado) para segmentar a la población estudiantil e identificar arquetipos de aprendizaje automáticos, analizando si estos grupos mejoran la capacidad predictiva del sistema.
- Implementar modelos de predicción secuencial basados en Transformers (Aprendizaje Supervisado), aprovechando el mecanismo de Self-Attention para detectar dependencias a largo plazo en el curso y superar las limitaciones de memoria de las redes recurrentes (LSTM) en la detección temprana de riesgo.
- Desplegar una estrategia de Explicabilidad Dual (XAI) para mitigar el problema de la "Caja Negra":
 - Intrínseca: Visualización de Mapas de Atención para identificar en qué momentos del curso se focaliza el modelo.
 - Agnóstica: Aplicación de valores SHAP (SHapley Additive exPlanations) —un método basado en la teoría de juegos que permite descomponer la predicción de cualquier modelo para entender el peso de cada factor— con el fin de cuantificar la contribución individual y global de cada variable en la predicción final del éxito o riesgo del estudiante.

- Validar empíricamente la superioridad del modelo propuesto, comparando sus métricas de precisión (Accuracy, F1-Score) y su capacidad explicativa frente a las líneas base tradicionales (Ensembles como XGBoost y reducciones lineales con PCA).
- Diseñar un marco de "Explicaciones Accionables", traduciendo los hallazgos técnicos (pesos de atención y valores SHAP) en recomendaciones pedagógicas concretas que permitan al docente realizar intervenciones personalizadas.

3. Marco teórico y estado del arte

La construcción de un sistema predictivo robusto exige un sólido anclaje en la literatura científica y pedagógica actual. Este capítulo establece los cimientos teóricos de la Minería de Datos Educativos (EDM) y la Analítica de Aprendizaje (LA), revisando las investigaciones de los últimos cinco años que sustentan el paso hacia el Deep Learning.

A través de este análisis, se justifica la adopción de arquitecturas híbridas —basadas en Autoencoders y Transformers— como la solución óptima para capturar la complejidad dinámica del comportamiento estudiantil en entornos virtuales. Asimismo, se posiciona la Inteligencia Artificial Explicable (XAI) como el puente necesario para transformar estos modelos complejos en herramientas pedagógicas transparentes, permitiendo al docente realizar intervenciones fundamentadas y accionables.

3.1. Marco teórico

A continuación, se desarrolla el marco conceptual que define la Minería de Datos Educativos (EDM) y la Analítica de Aprendizaje (LA), para posteriormente analizar el estado del arte de los últimos cinco años. Esta revisión sistemática permite identificar las limitaciones de los métodos tradicionales y justifica la elección de arquitecturas híbridas y técnicas de explicabilidad (XAI) como respuesta a la complejidad del comportamiento estudiantil en plataformas como Moodle.

3.1.1. Minería de Datos Educativos (EDM) y Learning Analytics.

La minería de datos educativos (EDM, Educational Data Mining) y la analítica del aprendizaje (LA, Learning Analytics) son técnicas que utilizan el análisis de datos para mejorar los procesos educativos. Para ello extrae patrones y tendencias del rendimiento, el comportamiento y las interacciones de los estudiantes. Estas disciplinas optimizan los procesos de aprendizaje, identifican a los estudiantes con sus dificultades y logran personalizar las experiencias de aprendizaje (Instituto

Andaluz Interuniversitario en Ciencia de Datos e Inteligencia Computacional (DASCI), s.f.).

Para materializar estas promesas, es fundamental comprender el origen de la información. Como se detalla en la revisión fundamental sobre EDM publicada en la revista *Modelling*, el primer paso crítico en cualquier ciclo de minería de datos es la recolección de registros provenientes de los "entornos educativos" donde interactúan los estudiantes (Papadogiannis, Wallace, & Karountzou, 2024). Estos entornos, principalmente los Sistemas de Gestión de Aprendizaje (LMS), actúan como la fuente primaria que alimenta los algoritmos.

3.1.2. Heterogeneidad de datos y estándares

En la actualidad existen una gran diversidad de instituciones educativas con plataformas educativas en línea. Cada una de estas plataformas tiene su propio sistema de gestión del aprendizaje, es decir, su propia manera de recolectar datos de los estudiantes, de estructurar la información y de delimitar que tipo de datos tienen mayor relevancia sobre otros, así como la relación entre estos. Esta falta de estandarización en la captura y estructuración de los datos plantea un desafío metodológico significativo: la dificultad para replicar estudios y generalizar hallazgos entre distintos entornos tecnológicos (Gašević, Dawson, Rogers, & Gasevic, 2016).

A consecuencia de esta tesitura, se vuelve necesario establecer un marco de referencia común que garantice la validez externa y la comparabilidad de los resultados (benchmarking). En este contexto, el conjunto de datos OULAD (Open University Learning Analytics Dataset) se ha consolidado como uno de los estándares de facto en la investigación sobre Minería de Datos Educativos. Este dataset destaca no solo por su amplia adopción en la literatura científica, sino por su riqueza multidimensional y su riguroso proceso de tratamiento y certificación garantizando la anonimidad de los datos (Véase Ilustración 1)

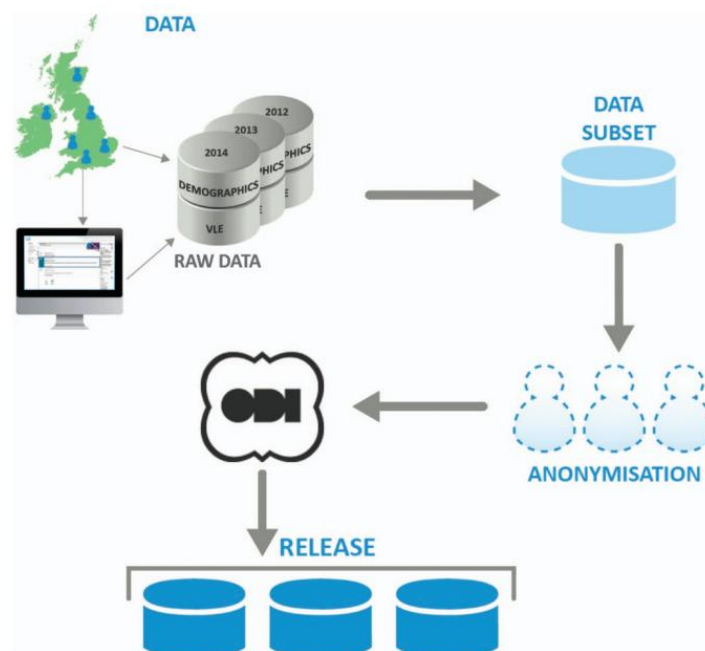


Ilustración 1 Proceso de preparación del dataset OULAD.

Referencia: (Kuzilek, Hlostá, & Zdrahal, 2017)

Integra datos demográficos, resultados de evaluación y registros detallados de interacción (logs) generados en el Entorno Virtual de Aprendizaje (VLE) de la Open University durante los cursos académicos 2013 y 2014. (Kuzilek, Hlostá, & Zdrahal, 2017)

Si bien han surgido conjuntos de datos más recientes en los últimos años, su idoneidad para la presente investigación es limitada. Gran parte de los datasets actuales (procedentes de competiciones tipo Kaggle o plataformas MOOC específicas) presentan inconvenientes críticos: o bien se trata de datos sintéticos generados artificialmente para sortear regulaciones de privacidad —lo que compromete la fidelidad del comportamiento humano real—, o bien carecen de la documentación académica rigurosa necesaria para sustentar una investigación formal. Asimismo, alternativas masivas como EdNet se centran excesivamente en el trazado de conocimiento (Knowledge Tracing) a nivel de ítem, alejándose de la estructura general de interacción típica de Moodle que este trabajo pretende

modelar. Por tanto, OULAD se mantiene como un recurso idóneo para validar modelos predictivos en entornos LMS tradicionales, cuya relevancia continúa siendo reconocida en los inventarios de datos de la literatura científica actual (Papadogiannis, Wallace, & Karountzou, 2024).

3.1.3. Deserción frente a bajo rendimiento

En el ámbito del *Learning Analytics* es crucial diferenciar entre la deserción (dropout) y el bajo rendimiento (failure).

La deserción se define como el cese de la actividad en la plataforma y la desvinculación formal del curso antes de su finalización. Este fenómeno suele estar asociado a factores motivacionales, gestión del tiempo o insatisfacción con el entorno. En el dataset OULAD, esto se representa mediante la etiqueta Withdrawn (retirada). Por el contrario, el bajo rendimiento se refiere a aquellos estudiantes que, completando el ciclo del curso y participando en las evaluaciones, no alcanzan los estándares mínimos de conocimiento, etiquetados como Fail.

Distinguir estos dos fenómenos es fundamental para la modelización computacional, ya que los patrones de interacción que preceden a un abandono (ej. disminución progresiva de accesos) difieren sustancialmente de los patrones de un estudiante que suspende (ej. accesos constantes, pero bajo desempeño en cuestionarios). Un sistema robusto debe ser capaz de diagnosticar ambos riesgos por separado.

3.1.4. Aprendizaje No Supervisado: Descubrimiento de Patrones (Clustering)

El Aprendizaje No Supervisado es una rama de la Inteligencia Artificial donde el modelo trabaja con datos no etiquetados, buscando estructuras ocultas o patrones intrínsecos en la información. En el contexto educativo, su aplicación principal es el clustering o agrupamiento.

A diferencia de la clasificación, donde se le dice al algoritmo "este es un buen alumno", el clustering agrupa a los estudiantes basándose únicamente en la similitud de sus comportamientos (ej. frecuencia de accesos, horarios de conexión, tipos de recursos visitados). Esto permite identificar arquetipos de estudiantes (perfiles) que no son evidentes a simple vista, como "el estudiante intensivo de fin de semana" o "el estudiante que solo lee foros, pero no participa". Algoritmos como K-Means (basado en centroides) o DBSCAN (basado en densidad) permiten segmentar la población estudiantil, proporcionando al docente una taxonomía del aula que facilita la personalización de la enseñanza.

3.1.5. Aprendizaje Supervisado y Modelos de Ensamble

El Aprendizaje Supervisado implica entrenar algoritmos con un conjunto de datos donde se conoce la respuesta correcta (la variable objetivo, en este caso, el resultado final del alumno), para que el modelo aprenda a predecir dicha variable en nuevos estudiantes basándose en su historial. En el contexto de la minería de datos educativos, la evolución de estos modelos ha pasado por tres fases diferenciadas:

- **Modelos Lineales:** Tradicionalmente, se empleaban técnicas como la Regresión Logística, insuficientes para capturar las relaciones complejas y no lineales del comportamiento estudiantil.
- **Modelos de Ensamble (Ensemble Learning):** Estudios comparativos recientes sobre el rendimiento estudiantil (Hasan, y otros, 2020) han validado que los algoritmos de ensamble, específicamente Random Forest, superan en precisión a las técnicas clásicas y a las redes neuronales simples cuando se trabaja con datos estructurados. Estos combinan múltiples "árboles de decisión" para manejar el desbalance de clases y determinar la importancia de las variables (Feature Importance). Sin embargo, tal como se evidencia en la metodología empleada por Hasan, y otros (2020), estos modelos dependen de características agregadas (ej. recuento total de visualizaciones), lo que reduce la compleja dinámica temporal del aprendizaje a valores estáticos acumulados, perdiendo el contexto secuencial.

- Deep Learning Secuencial y Transformers: Dado que el aprendizaje es un proceso dinámico que evoluciona en el tiempo, el estado del arte más avanzado se ha desplazado hacia redes neuronales capaces de procesar secuencias (Kusumawardani & Alfarozi, 2023).

Inicialmente, las redes recurrentes (RNN y LSTM) permitieron analizar la evolución temporal del alumno. Más recientemente, la arquitectura transformer ha revolucionado el campo. A diferencia de las anteriores, los Transformers utilizan mecanismos de Autoatención (Self-Attention). Esto permite al modelo ponderar la relevancia de cada evento del curso en relación con los demás, independientemente de la distancia temporal entre ellos. Gracias a esto, es posible detectar dependencias a largo plazo (por ejemplo, cómo el fracaso en una tarea de la semana 2 influye en el abandono en la semana 10) con una eficacia superior a cualquier otro método previo (Kusumawardani & Alfarozi, 2023).

3.1.6. Conclusiones

En síntesis, el Marco Teórico expuesto evidencia que tanto las técnicas no supervisadas como las supervisadas poseen fortalezas complementarias. Mientras que el clustering permite descubrir la estructura subyacente de los datos sin sesgos previos, los modelos supervisados capitalizan esa información para realizar predicciones concretas. Esta complementariedad teórica sugiere que la integración de ambas ramas —en un enfoque híbrido— constituye la arquitectura idónea para abordar la complejidad del comportamiento estudiantil en plataformas como Moodle, premisa que guiará la revisión de la literatura y la propuesta metodológica de este trabajo.

3.2. Estado del arte.

La literatura científica reciente en Minería de Datos Educativos ha evolucionado desde la aplicación aislada de algoritmos básicos hacia arquitecturas cada vez más complejas orientadas a la personalización. Para contextualizar la contribución de este TFM, esta sección analiza las investigaciones más relevantes de los últimos cinco años que han

utilizado el dataset OULAD. El análisis se estructura focalizándose en la evolución metodológica: partiendo de los enfoques puramente exploratorios (no supervisados) y predictivos (supervisados), hasta llegar a las propuestas híbridas más vanguardistas. Asimismo, se identifican las limitaciones recurrentes en los estudios actuales — específicamente en la reducción de dimensionalidad y la interpretabilidad— que justifican la necesidad de explorar técnicas no lineales de Deep Learning. Aprendizaje no supervisado en EDM

3.2.1. Aprendizaje no supervisado en EDM.

En el contexto específico del dataset OULAD, la investigación más reciente no solo valida la eficacia de las técnicas de clustering, sino que las posiciona como herramientas indispensables para la personalización educativa. Un ejemplo paradigmático es el estudio publicado por El Ghali, Atouf, El Guemmat, Broumi, & Talea (2025), quienes profundizaron en la segmentación estratégica de estudiantes utilizando este conjunto de datos. Su investigación contrastó algoritmos basados en vecindad (KNN) frente al clustering Jerárquico, empleando PCA (Análisis de Componentes Principales) para gestionar la alta dimensionalidad de las variables demográficas y de interacción.

Los resultados presentados por El Ghali et al. marcan el estándar del estado del arte actual: la aplicación del PCA no solo redujo el tiempo de cómputo en un 60%, sino que, combinado con clustering jerárquico, permitió alcanzar métricas de calidad excepcionales. Mediante esta metodología, lograron identificar cuatro arquetipos de estudiantes, revelando perfiles complejos como aquellos con "alta participación, pero bajo rendimiento" o patrones de compromiso erráticos (El Ghali, Atouf, El Guemmat, Broumi, & Talea, 2025).

Estos hallazgos son fundamentales para la presente investigación, pues confirman que OULAD contiene patrones latentes robustos. Sin embargo, la dependencia metodológica de este estudio reciente en el PCA sugiere que el campo sigue priorizando transformaciones lineales. Esto abre una oportunidad clara de mejora: si un modelo lineal como el PCA logra estos resultados, la aplicación de técnicas no lineales

(Autoencoders) podría capturar matices del comportamiento estudiantil que actualmente se pierden, siendo una de las hipótesis de este trabajo de fin de Máster.

3.2.2. Aprendizaje supervisado en EDM.

La evolución de los modelos predictivos sobre el dataset OULAD ha seguido una trayectoria clara: desde los clasificadores lineales básicos hacia arquitecturas de ensamble robustas y, más recientemente, hacia el aprendizaje profundo secuencial.

Si bien estudios iniciales establecieron líneas base con Regresión Logística y SVM, la literatura actual (2024-2025) posiciona a los algoritmos de Gradient Boosting como el estándar de facto para datos tabulares. Investigaciones recientes, como las de Mehmet Firat (2025), demuestran que XGBoost supera consistentemente a Random Forest, alcanzando precisiones notables. Sin embargo, estos autores coinciden en una limitación crítica: el rendimiento de los ensambles depende casi exclusivamente de una ingeniería de características manual, incapaz de capturar por sí sola la complejidad temporal del aprendizaje.

Para superar esta barrera, el estado del arte ha girado hacia el Deep Learning. Torkhani & Rezgui (2025) han validado el uso de redes LSTM (Long Short-Term Memory) sobre OULAD, logrando precisiones superiores al 83% gracias a su capacidad para procesar secuencias de datos. No obstante, las LSTM presentan limitaciones intrínsecas: su procesamiento secuencial impide la paralelización eficiente y sufren para retener dependencias a muy largo plazo (el problema del "olvido" en cursos largos).

En respuesta a estas ineficiencias, investigaciones emergentes han comenzado a señalar a la arquitectura Transformer como la evolución natural. Estudios recientes Kusumawardani & Alfarozi (2023) han implementado modelos de Transformer Encoder sobre el dataset OULAD, demostrando empíricamente que esta arquitectura supera a las redes recurrentes. Al reemplazar la memoria de estado oculta por capas de atención (Attention Layers), estos modelos no solo mejoran la precisión predictiva, sino que resuelven los problemas de coste computacional permitiendo una paralelización similar a las redes totalmente conectadas.

Más allá de la eficiencia, la ventaja decisiva de los Transformers para este TFM reside en su interpretabilidad intrínseca. A diferencia de las "cajas negras" tradicionales, los mecanismos de Autoatención (Self-Attention) generan matrices de pesos que indican explícitamente qué interacciones pasadas (ej. una tarea específica en la semana 2) influyeron más en la predicción final. Este TFM capitalizará dicha característica: se propone no solo utilizar transformers para maximizar la precisión, sino explotar sus mapas de atención —complementados con técnicas agnósticas como SHAP— para dotar al sistema híbrido de una capacidad explicativa (XAI) directa, permitiendo al docente entender el "porqué" del riesgo detectado sin sacrificar la potencia computacional.

3.2.3. El enfoque híbrido.

Los modelos híbridos representan la vanguardia en la minería de datos educativos, combinando técnicas no supervisadas y supervisadas en una arquitectura secuencial. A diferencia de los modelos monolíticos tradicionales, el enfoque híbrido segmenta primero la población mediante clustering para luego entrenar predictores especializados en cada perfil.

Esta estrategia ha sido validada empíricamente por Al-Tameemi et al. (2024) quienes demostraron que el uso de clústeres previos mejora significativamente la precisión de la clasificación en entornos educativos. No obstante, existe un patrón recurrente en la literatura de vanguardia, observado tanto en Al-Tameemi et al. como en El Ghali et al. (2025): la omnipresencia del PCA como técnica estándar de reducción de dimensionalidad, lo que implica una presuposición de que la relación entre los datos es lineal.

Dado que el aprendizaje humano es un proceso complejo y dinámico, esta asunción de linealidad podría estar simplificando la realidad de los datos. Por ello, este TFM propone una evolución metodológica sustituyendo la proyección lineal del PCA por una arquitectura de Autoencoders (Deep Learning). Esta aproximación busca capturar

las relaciones no lineales que el PCA omite, generando una representación latente más rica que potencie la precisión de los modelos predictivos posteriores.

3.2.4. Inteligencia Artificial Explicable (XAI) en Educación.

Tal como se adelantó al finalizar el análisis sobre los modelos supervisados y la arquitectura Transformer, la búsqueda de una mayor precisión predictiva conlleva la creación de "cajas negras", modelos con estructuras complejas que no son fácilmente interpretables (Khosravi, y otros, 2022). En el ámbito educativo, un modelo que predice el fracaso con un 90% de acierto pero que no explica las causas es pedagógicamente estéril, ya que impide al docente diseñar una intervención correctiva fundamentada. La literatura define esto como la necesidad de generar "explicaciones accionables" (actionable explanations), entendidas como datos que permiten establecer un procedimiento correctivo o bucle de retroalimentación para un conjunto de acciones (Khosravi, y otros, 2022).

Estudios recientes coinciden en que la adopción real de la Minería de Datos Educativos depende de la confianza (Trustworthiness) que el usuario final deposite en el sistema, siendo la explicabilidad el vehículo fundamental para incrementarla. Los modelos de Deep Learning avanzados, como las redes LSTM o los propios Autoencoders, operan mediante transformaciones no lineales complejas que oscurecen la relación directa entre la entrada y la salida; De hecho, la estructura de estas redes profundas dificulta su comprensión teórica incluso para expertos en la materia, debido a la geometría de sus espacios de alta dimensión (Sejnowski, 2020). Esta opacidad genera una barrera crítica: el docente no puede distinguir si una predicción de riesgo se debe a una falta de actividad, a un bajo rendimiento o a un patrón de comportamiento anómalo.

Para mitigar esta problemática sin renunciar a la potencia de los modelos no lineales, este TFM adopta una estrategia de explicabilidad dual, alineada con el marco de trabajo XAI-ED:

- **Explicabilidad Intrínseca (Attention Maps):** Aprovechando la arquitectura Transformer seleccionada, se extraerán y visualizarán las matrices

de pesos de atención (Self-Attention Weights). A diferencia de las redes recurrentes, estos mapas permiten observar directamente en qué momentos del curso o actividades específicas se "fijó" el modelo para determinar el riesgo de un estudiante. Este enfoque sigue las recomendaciones de complementar modelos complejos con componentes visuales interpretables para garantizar la transparencia (Khosravi, y otros, 2022).

- **Explicabilidad Agnóstica del Modelo (SHAP):** Dado que la arquitectura propuesta es híbrida, la interacción entre componentes puede ser compleja. Para garantizar una interpretación global, se utilizarán valores SHAP (Shapley Additive exPlanations). Esta técnica, fundamentada en la teoría de juegos, permite asignar una puntuación de contribución a cada variable, desvelando qué características empujaron la predicción hacia el éxito o el fracaso y distribuyendo equitativamente la importancia entre ellas (Lundberg & Lee, 2017).

3.2.5. Conclusiones

La revisión bibliográfica realizada, si bien no pretende abarcar la totalidad de la vasta producción científica en Minería de Datos Educativos, ha permitido identificar ciertas tendencias y patrones predominantes en los estudios recientes (2020-2025) sobre el dataset OULAD. A partir de la muestra seleccionada, se extraen tres conclusiones que fundamentan la propuesta experimental de este TFM:

1. **Exploración de la no-linealidad:** En los trabajos analizados (como El Ghali et al., 2025), se observa un uso frecuente de técnicas lineales como el PCA para la reducción de dimensionalidad. Si bien estas técnicas son efectivas, cabe la posibilidad de que la complejidad del comportamiento estudiantil contenga matices no lineales que se pierden en estas proyecciones. Por ello, resulta pertinente explorar si arquitecturas de Deep Learning, como los Autoencoders, pueden generar representaciones latentes más ricas y mejorar el rendimiento de los modelos posteriores en este contexto específico.

2. Evolución hacia modelos secuenciales: Aunque las redes LSTM han mostrado buenos resultados en la literatura consultada, los estudios más recientes comienzan a señalar las arquitecturas basadas en atención (Transformers) como una alternativa prometedora para capturar dependencias a largo plazo. Este trabajo busca contribuir a esta línea de investigación emergente, aplicando y validando la eficacia de los Transformers sobre los datos de la Open University.
3. La integración de la explicabilidad (XAI): Coincidiendo con la visión de autores como Sejnowski (2020) y Khosravi et al. (2022), se detecta una necesidad transversal de dotar de transparencia a los modelos complejos. Más allá de la precisión predictiva, este TFM asume la premisa de que la adopción real de estas herramientas depende de su interpretabilidad. Por tanto, se propone evaluar una estrategia de explicabilidad dual (SHAP y Attention Maps) para determinar si ofrece información pedagógicamente accionable al docente.

En definitiva, basándonos en el alcance de esta revisión, se constata que, si bien estas tecnologías han sido exploradas de manera aislada o parcial, no se ha hallado evidencia en la literatura reciente de una arquitectura que las orqueste simultáneamente sobre el dataset OULAD.

Por consiguiente, la combinación específica de Autoencoders (para la representación latente no lineal), Transformers (para el modelado secuencial) y XAI (para la explicabilidad dual) se presenta no solo como una arquitectura híbrida de alto interés experimental, sino como una propuesta novedosa que busca cubrir el vacío existente. Este enfoque pretende verificar si la integración sinérgica de estas tres técnicas avanzadas ofrece ventajas tangibles y superiores frente a la aplicación fragmentada o tradicional observada en el estado del arte.

4. Desarrollo del proyecto y resultados

Una vez sentado el marco teórico y analizados los antecedentes en la literatura científica, este capítulo constituye el núcleo práctico de la investigación. En las siguientes secciones se detalla el proceso técnico seguido para transformar los datos brutos del dataset OULAD en un sistema predictivo funcional. Se comienza describiendo la metodología de trabajo y el planteamiento del problema, para posteriormente desglosar las fases de ingeniería de características, el entrenamiento de la arquitectura híbrida y, finalmente, la discusión de los resultados obtenidos y su capacidad de explicación pedagógica.

4.1. Metodología.

La implementación de un sistema predictivo basado en *Deep Learning* dentro del contexto educativo requiere un enfoque estructurado que armonice los requisitos técnicos con las necesidades docentes. En este subapartado se detalla el marco metodológico adoptado, el cual permite transitar de manera sistemática desde la fase de comprensión de los datos brutos hasta la generación de acciones pedagógicas concretas y explicables.

Para garantizar la rigurosidad científica y la replicabilidad de la investigación, se describe a continuación el estándar procedimental seguido (CRISP-DM), la jerarquía de las tareas operativas diseñadas para este estudio junto con las tareas específicas a realizar y la hoja de ruta temporal que guía el desarrollo del proyecto.

4.1.1. Metodología CRISP-MD.

Para transformar los datos brutos de una plataforma educativa en conocimiento útil para un docente, no basta con aplicar algoritmos de forma aislada; es necesario seguir un proceso ordenado y probado. En este trabajo se ha optado por utilizar la metodología CRISP-DM (*Cross-Industry Standard Process for Data*

Mining), considerada el "estándar de oro" en los proyectos de ciencia de datos a nivel mundial (Wirth & Hipp, 2000).

A diferencia de un proceso lineal que tiene un principio y un fin rígidos, CRISP-DM es cíclico e iterativo: permite volver atrás y refinar lo aprendido en cada etapa para que el resultado final sea lo más preciso posible. Aunque nació en el ámbito industrial, la flexibilidad de CRISP-DM permite adaptarla perfectamente a las necesidades específicas de este Trabajo de Fin de Máster (TFM). Para este estudio, el ciclo de vida se ha personalizado en seis grandes bloques:

- Entender el problema (Comprensión del Negocio): El primer paso no es programar, sino entender la necesidad pedagógica: ¿Por qué los estudiantes abandonan los cursos? ¿Qué información necesita realmente un profesor para ayudarles a tiempo?
- Explorar los datos (EDA): Analizar la "huella digital" que dejan los alumnos en Moodle para entender qué tipo de información tenemos disponible y qué tan fiable es.
- Preparar la información: Limpiar y organizar esos datos para que las máquinas puedan "leerlos". En este TFM, esto implica convertir los clics diarios en secuencias que cuentan la historia del estudiante semana a semana.
- Construir los modelos (Modelado): Aquí es donde introducimos la potencia del Deep Learning. En lugar de usar modelos simples, entrenamos sistemas complejos (*Autoencoders y Transformers*) capaces de aprender patrones de comportamiento que no son obvios a simple vista.
- Verificar y Explicar (Evaluación y XAI): No basta con que el modelo acierte; necesitamos saber por qué lo hace. En esta fase integramos la Inteligencia Artificial Explicable (XAI) para traducir las decisiones matemáticas de la máquina en explicaciones claras para el docente.
- Aplicación de resultados (Despliegue): El paso final consiste en convertir esas predicciones en recomendaciones pedagógicas que puedan ser utilizadas en un entorno real de enseñanza.

Definida la metodología de trabajo, el siguiente paso consiste en delimitar el alcance del problema técnico y pedagógico. No basta con saber 'cómo' vamos a trabajar (CRISP-DM), sino que debemos precisar 'qué' estamos intentando resolver y cuáles son los hitos específicos que permitirán alcanzar el éxito del modelo.

4.1.2. Definición de tareas.

Para alcanzar los objetivos propuestos, el proyecto se ha desglosado en una serie de hitos técnicos y operativos. Estas tareas no solo cubren el desarrollo del código, sino también la fundamentación teórica y la transferencia de resultados al ámbito pedagógico:

Fase I: Fundamentación y contextualización

- Comprensión del ecosistema (Business Understanding): Análisis del dominio de la Minería de Datos Educativos (EDM) y la Analítica de Aprendizaje para alinear el modelo con las necesidades docentes.
- Definición de objetivos: Formalización del problema a resolver, estableciendo las métricas de éxito y los límites del sistema predictivo.
- Investigación del estado del arte: Revisión sistemática de literatura científica de los últimos cinco años para identificar arquitecturas de vanguardia y evitar redundancias en la investigación.

Fase II: Gestión y Exploración de Datos

- Adquisición de la fuente de datos: Selección y descarga del conjunto de datos OULAD, por su representatividad en interacciones tipo Moodle.
- Pre-procesamiento y Limpieza: Tratamiento de valores nulos, normalización de variables y filtrado de registros inconsistentes para asegurar la calidad de la entrada.
- Análisis Exploratorio de Datos (EDA): Identificación de patrones visuales, correlaciones y desequilibrios de clase (ej. tasa de abandono frente a suspensos).

Fase III: Ingeniería de Características y Modelado Híbrido.

- Ingeniería de Características I (Transformación Temporal): Conversión de los *logs* de interacción brutos en secuencias temporales estructuradas por semanas de curso.
- Modelado No Supervisado (Clustering): Implementación de algoritmos para descubrir arquetipos de estudiantes de forma automática.
- Ingeniería de Características II (Enriquecimiento): Integración de las etiquetas de los clústeres en el conjunto de entrenamiento para potenciar la capacidad de discriminación del modelo.
- Desarrollo del Modelo Supervisado: Diseño y entrenamiento de una red neuronal basada en la arquitectura Transformer, optimizada para capturar dependencias secuenciales a largo plazo.

Fase IV: Evaluación y Explicabilidad

- Evaluación del desempeño: Cálculo de métricas de rendimiento (Accuracy, F1-Score, Precisión) para validar la robustez de las predicciones.
- Implementación de Inteligencia Artificial Explicable (XAI): Aplicación de técnicas de interpretabilidad para desglosar el peso de cada variable y visualizar los mapas de atención del modelo.
- Acciones Pedagógicas (Transferencia): Traducción de los hallazgos técnicos en recomendaciones concretas para la intervención docente personalizada.

4.1.3. Planificación.

A continuación, se plasman en un diagrama de tipo Gantt la hoja de ruta cronológica y la distribución de esfuerzos diseñada para este trabajo. Este cronograma organiza temporalmente las cuatro fases operativas descritas anteriormente, estableciendo una secuencia lógica que garantiza el cumplimiento de los objetivos técnicos y pedagógicos dentro del plazo académico establecido.

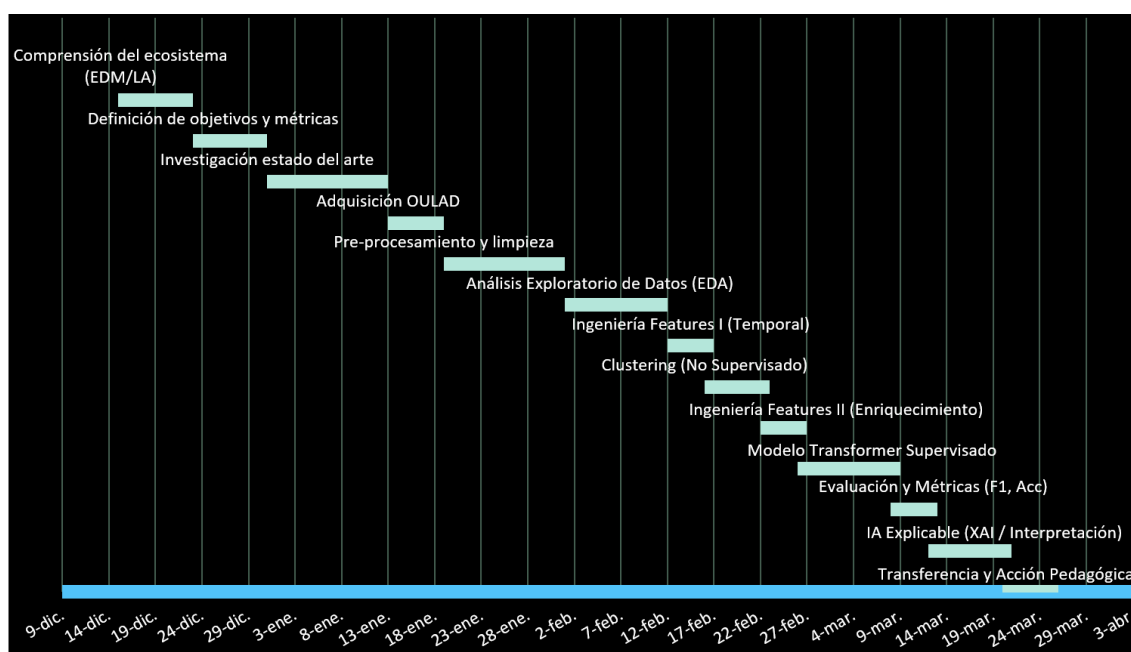


Ilustración 2 Cronograma de tareas definidas. Elaboración propia

5. Referencias

- Al-Tameemi, G., Xue, J., Ali, I. H., & Ajit, S. (2024). A Hybrid Machine Learning Approach for Predicting Student Performance Using Multi-class Educational Datasets. *Procedia Computer Science*, 238, 888-895. doi:10.1016/j.procs.2024.06.108
- El Ghali, M., Atouf, I., El Guemmat, K., Broumi, S., & Talea, M. (2025). ENHANCING E-LEARNING THROUGH STRATEGIC STUDENT SEGMENTATION: INSIGHTS FROM THE OULAD DATABASE . *Theoretical and Applied Information Technology*, 103(4), 1290-1299.
- Firat, M. (2025). Comparative Analysis of Random Forest vs XGBoost Machine Learning Algorithms for Predicting ODL Student Success.
- Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not ignore instructional conditions. *Computers & Education*, 53-66.
- Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K. U., & Sattar, M. U. (2020). Predicting Student Performance in Higher Educational Institutions Using Video Learning Analytics and Data Mining Techniques. *Applied Sciences*, 10(11), 3894. doi:10.3390/app10113894
- Instituto Andaluz Interuniversitario en Ciencia de Datos e Inteligencia Computacional (DASCI). (s.f.). *DASCI*. Recuperado el 2026, de Minería de datos educativos y análisis del aprendizaje: <https://dasci.es/linea-investigacion/mineria-de-datos-educativos/>
- Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., . . . Gašević, D. (2022). Computers and Education: Artificial Intelligence. *Computers and Education: Artificial Intelligence*, 3. Obtenido de <https://doi.org/10.1016/j.caeai.2022.100074>
- Kusumawardani, S. S., & Alfarozi, S. A. (2023). Transformer Encoder Model for Sequential Prediction of Student Performance Based on Their Log Activities. 11, págs. 18960-18971. IEEE. doi:10.1109/ACCESS.2023.3246122
- Kuzilek, J., Hlostá, M., & Zdrahal, Z. (2017). Open University Learning Analytics dataset. *Scientific Data*, 4(170171).
- Papadogiannis, I., Wallace, M., & Karountzou, G. (2024). Educational Data Mining: A Foundational Overview. *Encyclopedia*, 4(4), 1644-1664.

Sejnowski, T. J. (2020). The unreasonable effectiveness of deep learning in artificial intelligence. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1907373117

Torkhani, W., & Rezgui, K. (2025). OULAD MOOC Student Performance. (págs. 228-241). Atlantis Press. doi:10.2991/978-94-6463-654-3_18

Apéndice I

El apéndice es un adjunto al documento académico de autoría propia. No es un documento independiente, pues no se entendería si no es en relación con el resto del trabajo. Contiene información que complementa o aclara la tesis y que se considera que es demasiado larga o detallada para incluirse en el texto principal. Dicha información podría incluir gráficos o tablas, listas de datos sin procesar, etc.

Anexos I

Los anexos también contienen información adicional que se considera relevante para justificar las conclusiones del trabajo, pero, por lo general, el autor de contenido del anexo es distinto al autor del trabajo. Suele ser un documento independiente del trabajo. Pueden ser tablas de datos, imágenes, etc. Es necesario incluir las referencias de los documentos de donde procedan.