

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



viu

**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.....	7
1. Introducción.....	8
2. Objetivos.....	11
2.1. Objetivo General	11
2.2. Objetivos Específicos	11
3. Marco teórico y estado del arte.....	14
3.1. Marco teórico.....	14
3.1.1. Minería de Datos Educativos (EDM) y Learning Analytics.....	14
3.1.2. Heterogeneidad de datos y estándares	15
3.1.3. Deserción frente a bajo rendimiento	17
3.1.4. Aprendizaje No Supervisado: Descubrimiento de Patrones (Clustering)	17
3.1.5. Aprendizaje Supervisado y Modelos de Ensamble	18
3.1.6. Síntesis del marco teórico.....	19
3.2. Estado del arte.	20
3.2.1. Aprendizaje no supervisado en EDM.....	20
3.2.2. Aprendizaje supervisado en EDM.	21
3.2.3. El enfoque híbrido.....	23
3.2.4. Inteligencia Artificial Explicable (XAI) en Educación.	23
3.2.5. Síntesis del estado del arte	25
4. Desarrollo del proyecto y resultados.....	27
4.1. Planteamiento del problema.	27
4.2. Solución propuesta.....	28
4.3. Metodología.....	30

4.3.1.	Metodología CRISP-DM	30
4.3.2.	Definición de tareas.....	32
4.3.3.	Planificación.	34
4.4.	Desarrollo del proyecto.	35
4.4.1.	Comprendión del negocio.....	35
4.4.2.	Arquitectura y estructura del proyecto.	38
4.4.3.	Análisis exploratorio de datos (EDA).....	38
4.4.4.	Pre-procesamiento de datos.....	41
4.4.5.	Ingeniería de características I:.....	41
4.4.6.	Modelado I: Autoencoder	41
4.4.7.	Modelado 2: Clustering con Autoencoder y PCA	41
4.4.8.	Infraestructura de Ingeniería: Dockerización y Pipelines Automatizados	
	42	
5.	Resultados	43
6.	Conclusiones y trabajos futuros.....	43
8.	Referencias.....	44
	Apéndice I: Repositorio de Código y Reproducibilidad.....	46
	Anexos I.....	49



Índice de ilustraciones

Ilustración 1 Proceso de preparación del dataset OULAD. Referencia: (Kuzilek, Hlostá, & Zdrahal, 2017).....	16
Ilustración 2 Cronograma de tareas definidas. Elaboración propia.....	34



Índice de tablas

Tabla 1 Descripción de datos de estudiantes	37
Tabla 2 Descripción de entrega de actividades	37
Tabla 3 Descripción datos de interacción de estudiantes	38
Tabla 4 Ficha técnica de repositorio código fuente	46

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.

Tras una fase de ingeniería de características a partir de estos datos para transformar los logs en métricas de comportamiento, se aplicaron técnicas de aprendizaje no supervisado (agrupamiento) para segmentar a la población estudiantil e identificar arquetipos de aprendizaje y patrones de conducta. Estos hallazgos se incorporaron posteriormente a modelos de aprendizaje supervisado orientados a predecir el rendimiento final del estudiante (pass, fail, withdraw, distinction). Finalmente, se integraron técnicas de Inteligencia Artificial Explicable (XAI) para identificar las variables con mayor influencia en el riesgo académico, proporcionando información interpretable y accionable para apoyar intervenciones docentes personalizadas.

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

Abstract

Nowadays, Learning Management Systems (LMS) have become the fundamental infrastructure of the digital educational environment. Although these platforms have enabled the generation of a massive volume of data, a critical gap remains between data collection and its pedagogical use. In practice, LMS often operate as “black boxes”, storing records that teachers cannot interpret, which prevents the early detection of students at risk of failure.

This Master’s Thesis (TFM) addresses this problem through the development and validation of a hybrid Educational Data Mining (EDM) model. The research is based on the exploitation of the OULAD dataset (Open University Learning Analytics Dataset), one of the most comprehensive reference repositories worldwide, which integrates demographic information, assessment results, and detailed interaction logs from thousands of students across various university courses.

After a feature engineering phase to transform logs into behavioural metrics, unsupervised learning techniques (clustering) were applied to segment the student population and identify learning archetypes and behavioural patterns. These findings were subsequently incorporated into supervised learning models aimed at predicting students’ final performance (pass, fail, withdraw, distinction). Finally, Explainable Artificial Intelligence (XAI) techniques were integrated to identify the variables with the greatest influence on academic risk, providing interpretable and actionable information to support personalised teaching 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 uno de los LMS más extendidos en el ámbito educativo 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 (Gamage, Ayres, & Behrend, 2022). No obstante, a pesar de su uso extendido, el potencial de Moodle para la analítica de aprendizaje sigue estando infroutilizado, lo que hace necesario el diseño de marcos predictivos que aprovechen su analítica interna para identificar dificultades de aprendizaje (Soepriyanto, Nugroho, Nahri, Kesuma, & Setiasih, 2025).

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. Investigaciones recientes subrayan que estos *logs* por defecto de Moodle son fundamentales para analizar el comportamiento y los hábitos del estudiante, permitiendo modelar desde la retención hasta la carga de trabajo y los estilos de aprendizaje (Soepriyanto, Nugroho, Nahri, Kesuma, & Setiasih, 2025).

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, donde la complejidad de las estructuras de datos impide una interpretación directa por parte del profesorado (Khosravi, y otros, 2022).

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 aplicada a datos de interacción generados en un LMS, utilizando el dataset OULAD, que integre aprendizaje no supervisado (autoencoders y clustering) para construir representaciones latentes y perfiles de estudiantes, y redes Transformer para modelar la evolución temporal de dicha interacción, con el fin de predecir de forma temprana el riesgo de fracaso y abandono académico, incorporando además técnicas de inteligencia artificial explicable (XAI) para interpretar y justificar las predicciones del modelo.

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 modelar la evolución de la actividad del estudiante en el LMS y sus patrones de interacción a lo largo del curso, más allá de métricas estáticas acumuladas. Esta representación temporal resulta especialmente adecuada para su tratamiento mediante redes neuronales de tipo transformer y se alinea con propuestas recientes que destacan la utilidad de los “default logs” de Moodle para analizar el comportamiento del estudiante (Soepriyanto, Nugroho, Nahri, Kesuma, & Setiasih, 2025).
- 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 diferentes técnicas de aprendizaje no supervisado mediante algoritmos de clustering sobre distintas representaciones comprimidas de los datos —espacio latente aprendido por autoencoder y espacio reducido mediante PCA— con el fin de segmentar la población estudiantil, identificar arquetipos de aprendizaje emergentes y comparar hasta qué punto cada representación preserva o resalta estructura útil.
- 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.
- Comparar empíricamente el modelo transformers propuesto con líneas base tradicionales (Ensembles como XGBoost y reducciones lineales con PCA), analizando sus métricas de precisión (Accuracy, F1-Score) y el grado de explicabilidad, identificando en qué condiciones el enfoque propuesto aporta ventajas o limitaciones.
- 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, técnicas de clustering 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 se extraen 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 qué 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 situación, 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 un conjunto de datos de referencia ampliamente utilizado 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)

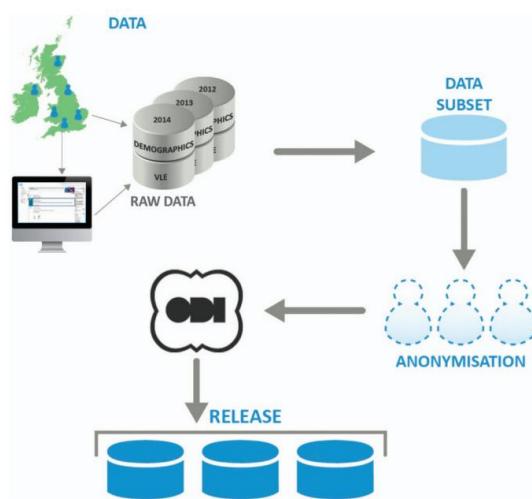


Ilustración 1 Proceso de preparación del dataset OULAD.

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 & Alfarizi, 2023).

Inicialmente, las redes recurrentes (RNN y LSTM) permitieron analizar la evolución temporal del alumno. Más recientemente, la arquitectura transformer ha ganado protagonismo en el modelado secuencial. A diferencia de las anteriores, los transformers se basan en mecanismos de autoatención (self-attention), que permiten ponderar la



relevancia de cada evento del curso en relación con los demás, independientemente de la distancia temporal entre ellos. De este modo, es posible capturar dependencias a largo plazo (por ejemplo, cómo el desempeño en una tarea temprana puede influir en el abandono en semanas posteriores) (Kusumawardani & Alfarizi, 2023; Kuzilek, Hlostá, & Zdrahal, 2017).

3.1.6. Síntesis del marco teórico

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.

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 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 relevantes para la presente investigación, ya que sugieren que el dataset OULAD contiene estructura y regularidades aprovechables. No obstante, el hecho de que el estudio reciente se apoye metodológicamente en PCA también refleja que el campo tiende a priorizar transformaciones lineales por su simplicidad e interpretabilidad. En este contexto, se propone explorar el uso de técnicas no lineales (autoencoders) como una vía complementaria para evaluar si pueden representar aspectos del comportamiento estudiantil que PCA no capture de forma directa. Esta propuesta se plantea de manera exploratoria, sin asumir a priori una mejora garantizada, y su utilidad se determinará empíricamente a partir de comparativas y métricas en las fases posteriores del trabajo.

3.2.2. Aprendizaje supervisado en EDM.

La evolución de los modelos predictivos aplicados a OULAD muestra una transición desde clasificadores lineales hacia métodos de ensamble y, más recientemente, hacia enfoques secuenciales de aprendizaje profundo. Este desplazamiento es coherente con la propia naturaleza del dataset OULAD, que combina variables demográficas, evaluaciones y trazas de interacción en el VLE para más de 32.000 estudiantes, y cuyo uso extendido facilita la comparación metodológica entre estudios (Kuzilek, Hlostá, & Zdrahal, 2017)

En tareas con variables tabulares agregadas, algoritmos de gradient boosting como XGBoost se emplean con frecuencia como línea base competitiva; por ejemplo, Firat (2025) reporta mejores resultados de XGBoost frente a Random Forest en su configuración experimental, destacando además la dependencia de estos modelos respecto a la ingeniería manual de características. En este sentido, parte de la literatura enfatiza que la ganancia de rendimiento en ensambles se apoya en “cómo” se agregan los logs (por semana/día, por tipo de actividad, por curso) y en la selección de variables, lo que puede limitar la captura de patrones temporales finos cuando se reduce la secuencia a totales acumulados.

Para incorporar la dimensión temporal de los registros del LMS, trabajos recientes han evaluado modelos secuenciales: Torkhani & Rezgui (2025) obtienen alrededor de un

83% de accuracy con LSTM, superando a varios enfoques clásicos en su comparación. Este resultado es consistente con la línea de trabajos que tratan los clics/interacciones como series temporales (p. ej., agregación semanal) para anticipar riesgo académico antes de finalizar el curso, priorizando el “cuándo” además del “cuánto” interactúa el estudiante. Además, otros enfoques recientes amplían el foco a escenarios multiclasificación (pass, fail, withdrawn, distinction) y muestran que redes neuronales pueden mejorar la predicción temprana cuando integran diversas fuentes (demografía, evaluaciones y clickstream) con preprocesado y feature engineering más exhaustivos (Junejo, y otros, 2024).

En esta misma línea, estudios como Kusumawardani & Alfarizi (2023) exploran transformer encoders sobre OULAD para predicción secuencial (diaria/semanal), reportando resultados competitivos y ventajas prácticas asociadas a la atención y la parallelización del entrenamiento. De forma complementaria, investigaciones recientes sobre OULAD centradas específicamente en abandono (withdrawal) también han propuesto variantes transformer y reportan un rendimiento alto y estable en predicción temprana cuando se usan secuencias semanales de actividad (Luo, Yu, Xie, Wu, & Xu, 2024). Este tipo de evidencias refuerza la idea de que, cuando se preserva la estructura secuencial de los logs, la autoatención puede competir (y en algunos diseños superar) a enfoques recurrentes en tareas de detección temprana de riesgo.

Más allá de la eficiencia, una posible ventaja de los transformers para este TFM es su utilidad interpretativa como señal complementaria: los mecanismos de autoatención generan pesos que permiten analizar qué partes de la secuencia (por ejemplo, interacciones en semanas concretas) han sido más relevantes para la predicción. No obstante, la literatura suele recomendar tratar la atención como una señal interpretativa “auxiliar” y complementarla con métodos de explicabilidad agnósticos al modelo para obtener justificaciones más robustas. En consecuencia, este TFM propone no solo evaluar modelos transformer en términos de rendimiento, sino también analizar patrones de atención y complementarlos con técnicas agnósticas como SHAP para reforzar la explicabilidad (XAI) del sistema híbrido, facilitando que el docente comprenda el “porqué” del riesgo detectado.

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, la asunción de linealidad puede resultar limitante en ciertos contextos y simplificar relaciones potencialmente más ricas en los datos. Por ello, este TFM plantea complementar el enfoque basado en PCA con una exploración de arquitecturas de autoencoders (Deep Learning) como alternativa no lineal de reducción de dimensionalidad. El objetivo es evaluar empíricamente si estas representaciones latentes son capaces de capturar dependencias que PCA no modela de forma directa y, en caso de aportar valor, utilizarlas como base para 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. Síntesis del estado del arte

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 matrices 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

Tras haber establecido los cimientos conceptuales en el marco teórico y analizado el estado del arte, este capítulo constituye el núcleo empírico de la investigación. Su propósito es detallar el proceso de ingeniería y modelado seguido para transformar los registros brutos del conjunto de datos OULAD en un sistema predictivo robusto y funcional.

A lo largo de las siguientes secciones, se describe la implementación técnica de una arquitectura híbrida que combina el aprendizaje de representaciones (Autoencoders) con el modelado secuencial (Transformers). El capítulo se estructura partiendo de la formalización del planteamiento del problema y la definición de la metodología (bajo el estándar CRISP-DM), para posteriormente abordar las fases de preprocesamiento, entrenamiento de los modelos y, finalmente, la discusión crítica de los resultados obtenidos, poniendo especial énfasis en su capacidad de explicación pedagógica (XAI) para el entorno docente.

4.1. Planteamiento del problema.

En la actualidad, la digitalización educativa ha convertido a los Sistemas de Gestión del Aprendizaje (LMS), especialmente Moodle, en infraestructuras críticas que generan un volumen masivo de datos. Cada interacción del estudiante genera una "huella digital" detallada en los archivos de registro o logs (Soepriyanto, Nugroho, Nahri, Kesuma, & Setiasih, 2025).

A pesar de esta abundancia de información, existe una brecha crítica entre la recolección de datos y su aprovechamiento pedagógico real. El problema se desglosa en tres ejes fundamentales:

- Opacidad del sistema: Los LMS operan frecuentemente como "cajas negras" donde los registros se almacenan sin ser interpretados (Khosravi, y otros, 2022).

- Detección tardía: Los docentes carecen de herramientas para identificar de forma proactiva a los estudiantes en riesgo de fracaso (*Fail*) o abandono (*Withdrawn*), detectando las dificultades cuando ya es demasiado tarde para intervenir.
- Falta de interpretabilidad: Los modelos predictivos avanzados suelen ser complejos y no ofrecen explicaciones claras sobre el "porqué" de una predicción, lo que los hace pedagógicamente estériles para el profesorado.
- Limitación en el modelado temporal: Gran parte de los enfoques actuales analizan el comportamiento del estudiante como una métrica estática acumulada (ej. número total de accesos), ignorando la naturaleza secuencial y evolutiva del aprendizaje. Esta simplificación impide detectar patrones complejos de cambio de comportamiento a lo largo de las semanas, que son indicadores críticos de riesgo.

Se requiere, por tanto, el desarrollo de un sistema inteligente que no solo procese masivamente estos registros para predecir el rendimiento final, sino que lo haga de manera explicable. Es imperativo transformar los datos brutos en "conocimiento accionable" que permita realizar intervenciones personalizadas y fundamentadas en patrones de comportamiento reales (Khosravi, y otros, 2022).

4.2. Solución propuesta.

Como respuesta a la problemática descrita, donde la infraestructura tecnológica supera a menudo la capacidad de interpretación pedagógica, este trabajo de fin de máster propone el diseño y validación de una arquitectura híbrida de inteligencia artificial. Esta propuesta busca cerrar la brecha entre la recolección masiva de registros (logs) y la generación de "insights" educativos, alineándose con la necesidad actual de marcos de trabajo que transformen datos brutos en alertas tempranas de riesgo. La solución se aleja de los enfoques tradicionales para adoptar un modelo capaz de procesar la complejidad secuencial del aprendizaje, integrando técnicas avanzadas de minería de datos educativos y Deep Learning.

La propuesta técnica se fundamenta en la integración sinérgica de tres componentes clave, diseñados para abordar cada uno de los ejes del problema identificado:

- Representación Latente y Segmentación (Abordando la complejidad de los datos) Para superar la heterogeneidad y el ruido inherente a los registros masivos de interacción en plataformas como Moodle, se propone utilizar técnicas de aprendizaje no supervisado y reducción de dimensionalidad. A diferencia de la ingeniería de características manual tradicional, el uso de arquitecturas de Deep Learning permite aprender representaciones complejas de los datos. Esto facilita la identificación de arquetipos de aprendizaje y la segmentación estratégica de los estudiantes, un paso fundamental para mejorar la personalización en entornos de e-learning (El Ghali, Atouf, El Guemmat, Broumi, & Talea, 2025). Asimismo, el enfoque híbrido propuesto busca potenciar la precisión predictiva al combinar estas representaciones con clasificadores robustos, una estrategia validada en estudios recientes sobre conjuntos de datos educativos multiclasificación (Al-Tameemi, Xue, Ali, & Ajit, 2024).
- Modelado Secuencial con Transformers (Abordando la detección tardía) Dado que el aprendizaje es un proceso evolutivo y no una métrica estática, la solución incorpora una arquitectura basada en Transformers y mecanismos de Autoatención (Self-Attention). Investigaciones recientes han demostrado que los modelos transformer encoder superan a las redes recurrentes tradicionales al capturar dependencias secuenciales a largo plazo en los registros de actividad de los estudiantes (Kusumawardani & Alfarozi, 2023). El sistema no solo evaluará el estado actual del alumno, sino su trayectoria completa, permitiendo emitir predicciones de rendimiento más precisas en entornos masivos. Este enfoque habilita un sistema de alerta temprana (Early Warning System) antes de que el fracaso sea irreversible, cumpliendo con el objetivo de soporte proactivo validado en la literatura reciente (Soepriyanto, Nugroho, Nahri, Kesuma, & Setiasih, 2025).
- Módulo de Explicabilidad Dual (Abordando la falta de interpretabilidad) Para mitigar el efecto de "Caja Negra" y garantizar la confianza y adopción por parte del docente, la arquitectura incluye una capa de inteligencia artificial explicable

(XAI). La literatura subraya que la IA en educación no debe limitarse a la predicción, sino que debe ofrecer transparencia sobre las decisiones algorítmicas (Khosravi, y otros, 2022). Por ello, esta solución no entrega solo una probabilidad de éxito, sino que desglosa el "porqué" mediante dos estrategias complementarias:

- Valores SHAP (Explicabilidad Agnóstica): Para cuantificar la contribución de cada variable a la predicción de riesgo.
- Mapas de Atención (Explicabilidad Intrínseca): Para visualizar en qué momentos específicos del curso se detectaron las anomalías.

En síntesis, la solución propuesta transforma el LMS de un repositorio pasivo de datos a un sistema de soporte a la decisión pedagógica, capaz de generar las "explicaciones accionables" necesarias para intervenir eficazmente en el proceso educativo.

4.3. 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.

4.3.1. Metodología CRISP-DM.

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. Este enfoque de transformación de datos brutos en nodos de información sigue la tendencia actual de utilizar analíticas internas de los LMS para identificar dificultades de aprendizaje de manera escalable (Soepriyanto, Nugroho, Nahri, Kesuma, & Setiasih, 2025).
- 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.

Adicionalmente, cabe subrayar que para este TFM, este bloque no se ha concebido como una serie de pasos manuales, sino como un pipeline de datos automatizado. Esta decisión metodológica garantiza que cualquier transformación (normalización, imputación o ingeniería de variables) se aplique de forma estricta e idéntica a los conjuntos de entrenamiento, validación y test, eliminando el riesgo de “fuga de datos” y asegurando la reproducibilidad total del experimento.

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.

4.3.2. 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.

4.3.3. 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.

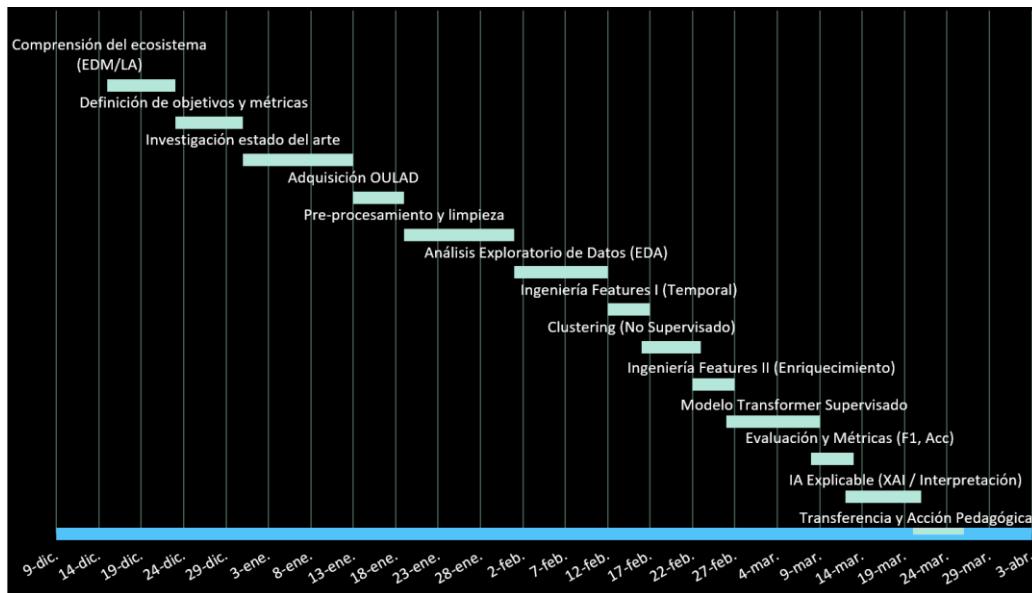


Ilustración 2 Cronograma de tareas definidas. Elaboración propia

4.4. Desarrollo del proyecto.

La implementación técnica de la arquitectura propuesta se ha llevado a cabo siguiendo los estándares de desarrollo de software y ciencia de datos. Cabe destacar que la totalidad del código fuente, los scripts de preprocesamiento y los cuadernos de experimentación se encuentran alojados en un repositorio público de GitHub ([Véase Apéndice I](#)). Esta decisión responde a un doble objetivo metodológico:

- Garantizar la trazabilidad y autoría: El historial de *commits* (registro de cambios) actúa como una prueba temporal veraz del progreso y la originalidad del trabajo realizado.
- Optimizar la memoria académica: Para evitar que este documento se extienda innecesariamente con bloques de código, este apartado se centra en describir la lógica, las decisiones arquitectónicas y los flujos de datos más relevantes. Para consultar la implementación detallada línea a línea, se remite al lector al repositorio mencionado o al anexo digital adjunto a esta entrega.

La estructura de este apartado sigue fielmente las fases del ciclo de vida CRISP-DM, detallando cómo se ha ejecutado cada etapa técnica:

4.4.1. Comprensión del negocio.

El proceso de comprensión del negocio se inició con una investigación exhaustiva sobre fuentes de datos que cumplieran con requisitos de fiabilidad, volumen y representatividad de un entorno LMS real. Tras evaluar diversas alternativas, se seleccionó el conjunto de datos OULAD (Open University Learning Analytics Dataset). La elección de este dataset se justifica por tres motivos críticos para la calidad de la investigación:

- Veracidad: Contiene datos reales de interacción estudiantil, no sintéticos.
- Anonimización: Cumple con los estándares éticos de privacidad mediante un riguroso proceso de desidentificación.

- Comparabilidad (Benchmarking): Al ser ampliamente utilizado en la literatura científica actual, permite contrastar directamente los resultados obtenidos por nuestra arquitectura frente a otros estudios previos, validando así las mejoras propuestas.

Cabe señalar que la fundamentación teórica del problema, así como el análisis profundo del estado del arte que contextualiza estos datos, se encuentran documentados en detalle en el Capítulo 3 de esta memoria.

Otro aspecto relevante para poder comprender la lógica de negocio del aprendizaje educativo en plataformas es entender los datos de OULAD observando cómo están estructurados y cómo se relacionan entre sí.

Para materializar este análisis estructural, se presenta a continuación el diccionario de datos detallado resultante de la etapa de preprocesamiento. Este desglose permite visualizar las tres dimensiones críticas que alimentarán el modelo híbrido propuesto: el perfil estático del estudiante, su rendimiento académico continuo y, fundamentalmente, la huella digital dinámica de su comportamiento en el entorno virtual.

- Tabla I: Dimensión del estudiante (students_processed.csv)

Esta tabla maestra consolida la información estática y demográfica, actuando como el perfil base de cada alumno matriculado.

Campo	Descripción
<i>code_module</i>	Identificador del curso (ej. AAA, BBB).
<i>code_presentation</i>	Semestre en el que se imparte (ej. 2013J, 2014B).
<i>id_student</i>	ID único anonimizado del estudiante.
<i>gender</i>	Género del estudiante (M/F).
<i>region</i>	Región geográfica de residencia en UK.
<i>highest_education</i>	Nivel educativo previo (ej. A Level, HE Qualification).
<i>imd_band</i>	Índice de Privación Múltiple (nivel socioeconómico; % más bajo indica mayor pobreza).
<i>age_band</i>	Rango de edad (0-35, 35-55, 55<=).
<i>num_of_prev_attempts</i>	Número de veces que el estudiante ha intentado este módulo previamente.
<i>studied_credits</i>	Carga total de créditos que el estudiante cursa

	actualmente.
<i>disability</i>	Indicador de discapacidad declarada (Y/N).
<i>final_result</i>	Variable Objetivo: Pass, Distinction, Fail, Withdrawn.
<i>date_registration</i>	Día relativo al inicio del curso en que se formalizó la matrícula.
<i>date_unregistration</i>	Día relativo en que se produjo la baja (si aplica).

Tabla 1 Descripción de datos de estudiantes

- Tabla II: Dimensión de Rendimiento (assessments_processed.csv)

Contiene los resultados cuantitativos de los hitos evaluativos, permitiendo trazar la evolución académica del alumno.

Campo	Descripción
<i>id_assessment</i>	ID único de la prueba específica.
<i>id_student</i>	ID del estudiante.
<i>date_submitted</i>	Día relativo al inicio del curso en que se realizó la entrega.
<i>is_banked</i>	Indicador de si la nota ha sido convalidada de una presentación anterior.
<i>score</i>	Calificación obtenida (escala 0-100).
<i>assessment_type</i>	Tipología: TMA (Corrección por tutor), CMA (Corrección automática), Exam.
<i>date</i>	Fecha límite oficial de entrega (<i>Deadline</i>).
<i>weight</i>	Ponderación de la evaluación en la nota final del curso.

Tabla 2 Descripción de entrega de actividades

Tabla III: Dimensión de Comportamiento (interactions_processed.csv)

Representa el log de actividad diario en el Entorno Virtual de Aprendizaje (VLE).

Dada su granularidad, constituye la fuente principal para el modelado secuencial.

Campo	Descripción
<i>code_module</i>	Identificador del curso (ej. AAA, BBB).
<i>code_presentation</i>	Semestre en el que se imparte (ej. 2013J, 2014B).
<i>id_student</i>	ID único anonimizado del estudiante.
<i>gender</i>	Género del estudiante (M/F).
<i>region</i>	Región geográfica de residencia en UK.
<i>highest_education</i>	Nivel educativo previo (ej. A Level, HE Qualification).
<i>imd_band</i>	Índice de Privación Múltiple (nivel socioeconómico; % más bajo indica mayor pobreza).
<i>age_band</i>	Rango de edad (0-35, 35-55, 55<=).
<i>num_of_prev_attempts</i>	Número de veces que el estudiante ha intentado este módulo previamente.

<i>studied_credits</i>	Carga total de créditos que el estudiante cursa actualmente.
<i>disability</i>	Indicador de discapacidad declarada (Y/N).
<i>final_result</i>	Variable Objetivo: Pass, Distinction, Fail, Withdrawn.
<i>date_registration</i>	Día relativo al inicio del curso en que se formalizó la matrícula.
<i>date_unregistration</i>	Día relativo en que se produjo la baja (si aplica).

Tabla 3 Descripción datos de interacción de estudiantes

Una vez desglosada la arquitectura de la información y definidas las variables operativas de las tres dimensiones clave (perfil, rendimiento y comportamiento), se da por concluida la fase de definición estructural. Para validar la integridad de estos registros y asegurar que representan fielmente la realidad educativa que se pretende modelar, es imperativo someter dichos valores a un escrutinio estadístico detallado mediante un análisis exploratorio de los datos, proceso que se aborda en el siguiente epígrafe.

4.4.2. Arquitectura y estructura del proyecto.

4.4.3. Análisis exploratorio de datos (EDA).

A nivel técnico, la exploración de los datos (EDA) se materializó mediante scripts de análisis en un jupyter notebook, concretamente en el archivo *data_exploration.ipynb*. Este cuaderno, cuya ubicación exacta dentro de la estructura del proyecto se detalla en el Apéndice I, permite examinar las distribuciones de las variables y la correlación entre las interacciones y el rendimiento final.

Del análisis realizado en dicho cuaderno, se extrajeron tres conclusiones determinantes para la siguiente fase de ingeniería:

El Análisis Exploratorio de Datos permitió identificar patrones estructurales y anomalías en el conjunto *OULAD* que guiaron el diseño de la etapa de limpieza y

transformación de datos. A continuación se detallan los hallazgos principales y las acciones correctivas aplicadas en el módulo dataset.py:

1. Gestión de Valores Nulos en Variables Críticas

- Observación (EDA): Se detectó que la variable `imd_band` (Índice de Privación Múltiple), un indicador socioeconómico clave, presentaba aproximadamente 1.100 valores nulos.
- Decisión de Diseño: Dado que la ausencia de información socioeconómica podría correlacionarse con perfiles de alumnos específicos (ej. internacionales o datos protegidos), se desestimó la eliminación de estos registros para evitar sesgos de selección.
- Acción de Preprocesamiento: Se imputaron estos valores nulos mediante una nueva categoría explícita denominada

"Unknown"

. Esto permite al modelo interpretar la falta de dato como una característica informativa en sí misma, preservando la integridad del dataset.

2. Preservación de Historial de Interacciones (Imputación Temporal)

- Observación (EDA): Una pequeña fracción de la muestra (~45 estudiantes) carecía de `date_registration`.
- Decisión de Diseño: Eliminar estos registros supondría una pérdida injustificada de datos conductuales valiosos para el entrenamiento de redes neuronales secuenciales (Transformers/LSTMs).
- Acción de Preprocesamiento: Se implementó una estrategia de imputación estadística condicional, asignando a los valores faltantes la mediana de los días de registro, calculada específicamente para cada cohorte (`code_presentation`

). Esta técnica minimiza la distorsión de la distribución temporal original.

3. Consistencia en el Sistema de Evaluación

- Observación (EDA): Se identificaron inconsistencias en la tabla de evaluaciones, incluyendo registros sin puntuación (

score

) o con metadatos erróneos.

- Decisión de Diseño: Para garantizar la fiabilidad del entrenamiento supervisado, es imperativo que las etiquetas de rendimiento (notas) sean veraces.
- Acción de Preprocesamiento: Se aplicó un filtrado estricto eliminando únicamente aquellos registros de evaluación que carecían de nota final, asegurando así que todas las series temporales de rendimiento alimentadas al modelo estén completas y saneadas.

4. Normalización de Tipos de Datos

- Observación (EDA): Tras la carga inicial, múltiples variables numéricas (como días relativos y puntuaciones) fueron interpretadas erróneamente como objetos debido a la presencia de caracteres no estándar (ej.

?

).

- Acción de Preprocesamiento: Se estandarizó la conversión de tipos mediante

`pd.to_numeric`

con coerción de errores tras la limpieza de caracteres especiales. Esto garantiza que las matrices de características resultantes sean numéricamente operables para los algoritmos de Machine Learning.

4.4.4. Pre-procesamiento de datos.

4.4.5. Ingeniería de características I:

La arquitectura propuesta explota la capacidad de los Transformers para procesar secuencias multivariantes complejas, integrando señales heterogéneas (comportamiento de navegación + rendimiento académico) en un mismo espacio latente temporal.

4.4.6. Modelado I: Autoencoder

4.4.7. Modelado 2: Clustering con Autoencoder y PCA

Elegimos Gaussian Mixture Models (GMM) como técnica de segmentación por tres motivos principales. En primer lugar, su naturaleza probabilística permite asignar a cada estudiante una distribución de pertenencia a los distintos grupos, en lugar de imponer fronteras duras. Esta propiedad es especialmente adecuada en nuestro contexto, ya que buscamos representar perfiles de comportamiento derivados de embeddings (p. ej., basados en arquitecturas Transformer) donde es razonable asumir transiciones graduales entre patrones y, por tanto, incertidumbre en la asignación.

En segundo lugar, GMM ofrece un buen compromiso entre capacidad descriptiva y granularidad. Con $k=6$ se obtienen agrupaciones suficientemente finas como para capturar heterogeneidad relevante, sin caer en una segmentación excesivamente fragmentada. Esto permite superar la simplificación típica del dataset original (p. ej., withdraw / pass / fail), obteniendo perfiles más ricos e interpretables, útiles para el análisis y potencialmente para diseñar intervenciones educativas diferenciadas.

Por último, al comparar el comportamiento de GMM sobre la representación lineal (PCA) y sobre el espacio latente aprendido por el Autoencoder, las métricas y las visualizaciones sugieren que el espacio latente del Autoencoder produce una estructura más coherente y separable, además de perfiles con diferencias más consistentes en la

tasa de éxito. Por ello, adoptamos GMM sobre el espacio latente del Autoencoder como la configuración principal para el análisis posterior.

4.4.8. Infraestructura de Ingeniería: Dockerización y Pipelines Automatizados

Para garantizar que la arquitectura de Deep Learning sea agnóstica al hardware diferentes dispositivos y así garantizar su reproducibilidad se ha diseñado el proyecto bajo una infraestructura modular y reproducible basada en dos pilares:

Contenerización del entorno (Docker): Se ha implementado una arquitectura de contenedores que aísla las dependencias críticas (como las bibliotecas CUDA/cuDNN para aceleración por GPU):

- **Dockerfile:** Define una imagen base inmutable con la versión exacta de python 3.11. Incluye todas las capas de sistema necesarias para la aceleración por hardware y las bibliotecas de cálculo numérico (TensorFlow, Scikit-learn), garantizando que el modelo se comporte igual en un servidor local que en la nube, o en cualquier hardware.
- **Docker-Compose:** Actúa como el orquestador del entorno, configurando de forma automática el montaje de volúmenes de datos, la gestión de puertos y la inyección de variables de entorno, permitiendo levantar el "laboratorio experimental" completo con un solo comando.
- **Dev Containers:** Se ha integrado un estándar de desarrollo que permite vincular el IDE directamente con el contenedor. Esto asegura que el código escrito y el código ejecutado comparten siempre el mismo ecosistema, librerías y versiones de compiladores.

Orquestación del Pipeline (Makefile): A diferencia del enfoque tradicional basado exclusivamente en notebooks experimentales, este proyecto implementa una Pipeline-as-Code profesional. Se ha diseñado un archivo maestro de orquestación (Makefile) que automatiza el flujo de trabajo completo:

- Ingesta de datos (make data): Realiza la limpieza, el limpiado preventivo de directorios y la división del dataset (Split) de forma determinista.
- Entrenamiento y Latencia (make_prep_latent): Encadena de forma secuencial el entrenamiento del Autoencoder y la generación de embeddings, asegurando la integridad referencial entre los modelos y sus representaciones vectoriales.
- Reproducibilidad Total: Este sistema permite que cualquier investigador pueda replicar el experimento completo desde cero ejecutando un comando único, eliminando la intervención manual y los posibles sesgos humanos en el preprocesamiento de los datos.

5. Resultados

6. Conclusiones y trabajos futuros.

Hay una **zona gris** entre Fail y Withdrawn. Algunos alumnos que figuran como "Withdrawn" en realidad abandonaron *porque iban a suspender* (se retiraron antes del examen final para no manchar el expediente). Otros abandonaron por motivos **externos** (trabajo, salud, economía).

Tu modelo no puede distinguir eso porque el dataset no lo refleja. Pero es un **punto excelente para la sección de "Limitaciones" de tu TFM**. Los revisores valoran mucho que reconozcas estas fronteras difusas.

8. 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/>
- Junejo, N. U., Nawaz, M. W., Huang, Q., Dong, X., Wang, C., & Zheng, G. (2024). Accurate Multi-Category Student Performance Forecasting at Early Stages of Online Education Using Neural Networks. *arXiv preprint arXiv:2412.05938*, 1-27.
- 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., Hlosta, M., & Zdrahal, Z. (2017). Open University Learning Analytics dataset. *Scientific Data*, 4(170171).

Luo, H., Yu, Q., Xie, S., Wu, H., & Xu, K. (2024). GLU-Transformer for Predicting Withdrawal in VLE. *ICBAR '24: Proceedings of the 2024 4th International Conference on Big Data, Artificial Intelligence and Risk Management*, 835-840.

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

Soepriyanto, Y., Nugroho, R. P., Nahri, M. H., Kesuma, D. W., & Setiasih, M. (2025). From logs to insights: A comprehensive framework for data-driven learning insights. *Jurnal Inovasi Teknologi Pendidikan*, 12(1), 40-49. doi:<https://doi.org/10.21831/jitp.v12i1.77432>

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: Repositorio de Código y Reproducibilidad

Acceso al Repositorio Todo el código fuente, los notebooks de exploración (EDA) y los scripts de entrenamiento desarrollados para este Trabajo de Fin de Máster se encuentran alojados en un repositorio público de GitHub para garantizar la transparencia y reproducibilidad de los resultados.

Atributo	Detalle
URL del Repositorio	https://github.com/MichaeLaudrup/Artificial_Intelligence_Msc/tree/main/TFM_education_ai_analytics
Plataforma	GitHub
Lenguaje Principal	Python 3.12+
Librerías Clave	Pandas, TensorFlow/Keras, Scikit-learn, SHAP
Licencia	MIT License (Código abierto)
Código QR para acceso directo	

Tabla 4 Ficha técnica de repositorio código fuente

Estructura del Repositorio: El proyecto sigue una arquitectura modular inspirada en los estándares de la industria para ciencia de datos, lo que garantiza la separación de preocupaciones y facilita la reproducibilidad del experimento. A continuación, se detalla la jerarquía de directorios y la función de sus componentes principales:

root/: Directorio principal del proyecto que contiene los archivos de configuración y automatización.

- data/: Almacén central de datos, organizado por etapas de madurez:
 - raw/: Copias inalteradas de los datos originales (OULAD).
 - interim/: Datos transformados durante los procesos de limpieza intermedios.
 - processed/: Conjuntos de datos finales, anonimizados y listos para alimentar los modelos de IA.
 - external/: Datos complementarios de terceras fuentes utilizados en el análisis.
- notebooks/: Cuadernos interactivos (.ipynb) utilizados para la experimentación inicial.
 - data_exploration.ipynb: Documenta el análisis exploratorio (EDA), visualizaciones estadísticas y la validación de hipótesis iniciales.
- educational_ai_analytics/: Núcleo del código fuente desarrollado como un paquete modular de Python:
 - dataset.py: Funciones para la extracción y carga eficiente de los datos.
 - features.py: Implementación de la ingeniería de variables y transformadores.
 - modeling/: Scripts dedicados al entrenamiento, validación cruzada y serialización de los modelos.
 - plots.py: Motor de visualización personalizado para la generación de gráficas presentadas en este trabajo.
 - config.py: Definición de hiperparámetros, rutas del sistema y constantes globales.
- models/: Directorio destinado a almacenar los pesos de los modelos entrenados y los archivos de metadatos asociados.
- reports/: Resultados del análisis.
 - figures/: Gráficas y representaciones visuales exportadas en alta resolución para la memoria del TFM.
- references/: Documentación técnica, diccionarios de datos y manuales de las fuentes de información utilizadas.



- Makefile: Archivo de automatización que permite replicar todo el flujo de trabajo (limpieza, procesado y entrenamiento) mediante comandos simples (ej. make data, make models).
- pyproject.toml: Especificación de la versión de Python (3.12+) y lista exhaustiva de dependencias necesarias para recrear el entorno virtual.



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.