ENVIHEI

Título del reto de ABP:
La ingeniería de datos se une a la ciencia climática: Exploración de CMIP6 mediante el aprendizaje basado en problemas

Nombre del profesor responsable

Andrés Navarro Martínez de la Casa

Universidad / Departamento / Titulación

Universidad de León / Departamento de Química y Física Aplicadas / Grado en Ingeniería de Datos e Inteligencia Artificial

Curso / Asignatura en la que se ha implementado

4º curso / Proyectos en Ingeniería de Datos e Inteligencia Artificial

Periodo de implementación (fechas)

10 de septiembre de 2025 – 31 de octubre de 2025

Duración total (semanas/horas)

8 semanas / 30 horas

Número de estudiantes participantes

7 estudiantes

Todos los resultados desarrollados en el marco de ENVIHEI están disponibles bajo la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Cofinanciado por la Unión Europea. No obstante, las opiniones y los puntos de vista expresados son exclusivamente los del autor o autores y no reflejan necesariamente los de la Unión Europea o del Servicio Español para la Internacionalización de la Educación (SEPIE). Ni la Unión Europea ni el SEPIE pueden ser considerados responsables de ellos. [2024-1-ES01-KA220-HED-000246733]

1. Diseño del proyecto

1.1 Breve descripción del proyecto y del reto propuesto

La asignatura es una experiencia de aprendizaje basado en problemas (ABP) diseñada para estudiantes de ingeniería de datos. El reto principal consiste en aplicar el ciclo de vida completo de la ingeniería de datos – ingesta, limpieza, análisis y modelización – a problemas climáticos reales. Los estudiantes procesarán gigabytes de datos de modelos climáticos (CMIP6) y de reanálisis (ERA5) para responder a preguntas complejas sobre sostenibilidad, como la detección de anomalías climáticas, la clasificación de zonas bioclimáticas mediante aprendizaje automático no supervisado y la evaluación de la incertidumbre de las proyecciones climáticas futuras mediante conjuntos de múltiples modelos.

1.2 Resultados de aprendizaje esperados (competencias académicas y transversales)

Se espera que los estudiantes demuestren un dominio tanto de la ingeniería de datos avanzada como de sus aplicaciones a la ciencia climática. Los resultados de aprendizaje académicos se centran en el diseño y la implementación de flujos automatizados de ETL (extracción, transformación, carga) para datos geoespaciales de alta dimensionalidad. Los estudiantes aprenden a integrar las API de datos satelitales y modelizados (Copernicus CDS y ESGF) con formatos de almacenamiento eficientes. También aprenden a aplicar técnicas de aprendizaje automático no supervisado – específicamente la reducción de dimensionalidad (ACP) y el agrupamiento (K-Means) – para descubrir patrones espaciotemporales latentes, al tiempo que cuantifican rigurosamente la incertidumbre predictiva mediante el análisis estadístico de conjuntos de múltiples modelos (CMIP6).

De forma transversal, el proyecto fomenta la capacidad de gestionar la incertidumbre en los procesos de toma de decisiones, una competencia fundamental al trabajar con modelos climáticos probabilísticos. Los estudiantes también dominan los estándares de ciencia reproducible mediante el uso de sistemas de control de versiones y buenas prácticas de documentación. Por último, la asignatura desarrolla sólidas habilidades de narración basada en datos, que permiten a los estudiantes sintetizar conjuntos de datos complejos y multidimensionales en información visual útil para la acción que sirva de base para las decisiones de diversas partes interesadas.

1.3 Competencias en sostenibilidad que se pretende desarrollar

Este proyecto aborda directamente las metas de tres Objetivos de Desarrollo Sostenible (ODS), de acuerdo con las directrices de la Red de Soluciones para el Desarrollo Sostenible (SDSN/REDS):

- **ODS 13 - Acción por el clima (Objetivo principal)**: Formación técnica para analizar variables climáticas esenciales y comprender la base física del cambio climático utilizando datos observados y modelizados.
- **ODS 9 - Industria, innovación e infraestructura**: Fomento de la investigación científica y de la capacidad tecnológica mediante el uso de 'macrodatos' e inteligencia artificial aplicados a la sostenibilidad.
- **ODS 4 - Educación de calidad**: Adquisición de competencias digitales avanzadas y promoción del acceso público a la información climática (Datos abiertos/Ciencia abierta).

1.4 Enfoque de sostenibilidad y su integración en el proyecto

La sostenibilidad no es un añadido, sino el contenido central de la asignatura. Cada módulo técnico está vinculado a una cuestión de sostenibilidad: el cálculo de normales climatológicas para la adaptación climática y las infraestructuras

resilientes (ASS01), el uso de conjuntos de modelos para gestionar la incertidumbre en las predicciones (ASS02) y la redefinición de las zonas climáticas bajo escenarios de cambio global (ASS04). Estas conexiones obligan a los estudiantes a comprender la física del cambio climático para desarrollar soluciones de datos eficaces.

1.5 Metodologías y herramientas utilizadas (estructura del ABP, técnicas de facilitación, etc.)

El reto orientador. El proyecto se articula en torno a una pregunta central abierta que guía todo el proceso de aprendizaje: ¿Cómo podemos transformar datos climáticos en bruto, a escala de petabytes, en información útil para actuar en favor de la sostenibilidad en un mundo cambiante? Los estudiantes actúan como ingenieros de datos encargados de 'descodificar' el sistema climático para abordar este reto.

Estructura del ABP & Andamiaje. La metodología sigue un enfoque de andamiaje progresivo, pasando de la indagación guiada a la resolución autónoma de problemas a lo largo de tres fases (fundamentación, indagación y resolución):

- **Fase de fundamentación (El conjunto de herramientas)**. Se introduce a los estudiantes en el ecosistema (Linux/Conda) y aprenden a automatizar la ingesta de datos mediante API. El énfasis se sitúa en la competencia técnica.
- **Fase de indagación (La física de los datos)**. Los estudiantes abordan subproblemas específicos (normales climatológicas, anomalías y conjuntos de modelos), que requieren comprender los conceptos físicos subyacentes para escribir el código correcto. Esta fase conecta la estadística con la sostenibilidad.

- **Fase de resolución (El proyecto integrador)**. El proyecto final 'Sistema de clasificación climática (CCS)' elimina los apoyos iniciales. Los estudiantes deben integrar todos los módulos anteriores (obtención de datos, limpieza, análisis estadístico y aprendizaje automático) para proponer una solución novedosa al reto orientador.

Técnicas de facilitación:

- **Enseñanza justo a tiempo (JiTT)**: La JiTT introduce conceptos teóricos (p. ej., conjunto de modelos, anomalía, ACP) cuando los estudiantes los necesitan para resolver un problema de programación, lo que maximiza la pertinencia y la retención del aprendizaje.
- **Dojos de programación**: Sesiones prácticas en las que el profesor programa en directo para demostrar estrategias de depuración y buenas prácticas que los estudiantes después replican y amplían.
- **Revisiones de código entre pares (In situ)**: En lugar de la calificación tradicional, los estudiantes reciben retroalimentación mediante revisiones de código 'por encima del hombro'.

Herramientas & Ecosistema. La implementación técnica se basó en un conjunto de tecnologías de código abierto estándar en la industria.

- **Entorno**: Linux/WSL con Conda para la gestión de paquetes.
- **Bibliotecas principales**: Python (Xarray, Pandas, Matplotlib y Cartopy) para la manipulación y visualización de datos.
- **Herramientas específicas para el clima**: CDO (operadores de datos climáticos) para el procesamiento de alto rendimiento de NetCDF y GDAL para operaciones con datos ráster geoespaciales.
- **Acceso a los datos**: Obtención automatizada mediante la API de Copernicus CDS y los nodos de ESGF.

- **Colaboración**: Se utilizan Git y GitHub para el control de versiones y la entrega de proyectos.

1.6 Organización del proyecto (formación de grupos, cronograma, hitos)

8 semanas (30 horas presenciales + trabajo autónomo):

- **Semana 1**: Fundamentos, herramientas y adquisición de datos (API).
- **Semanas 2-4**: Estadística climática (normales climatológicas, anomalías y conjuntos de modelos). Tareas individuales.
- **Semanas 5-7**: Aprendizaje automático avanzado (Agrupamiento del CCS). Los estudiantes trabajaron en grupos pequeños, replicando un entorno de desarrollo profesional.
- **Semana 8**: Integración final y presentaciones orales.

1.7 Entregables solicitados a los estudiantes (tipo, formato, plazos)

Para cada módulo (ASS01-ASS04), se exigieron los siguientes elementos:

- **Código**: Scripts o cuadernos de Python limpios y documentados en el repositorio y/o en Moodle.
- **Informe técnico**: Documento PDF que incluya la metodología, las visualizaciones y la discusión de los resultados.
- **Visualizaciones**: Mapas o gráficos generados con Cartopy/Matplotlib para cada tarea.
- **Presentación final**: Defensa oral del proyecto CCS. Solo para el proyecto colaborativo (ASS04).

Plazos: Un entregable por semana (siete días después de la sesión), salvo en ASS04, que es un proyecto colaborativo de cuatro semanas.

1.8 Colaboradores externos participantes (mentores, empresas, expertos...) y su función

No participaron colaboradores externos en esta implementación.

2. Evaluación del proyecto

2.1 Criterios y métodos de evaluación utilizados

La evaluación fue continua y se basó en rúbricas específicas para cada tarea, con la siguiente ponderación habitual:

- **Corrección técnica y análisis de datos (40%)**: Precisión en los cálculos (p. ej., anomalías, medias estacionales).
- **Calidad del código y reproducibilidad (20%)**: Estructura, comentarios y uso eficiente de las bibliotecas.
- **Calidad de las visualizaciones (20%)**: Claridad comunicativa de los mapas y gráficos.
- **Informe y comunicación (20%)**: Capacidad para sintetizar e interpretar los resultados matemáticos.

Los detalles de las rúbricas específicas pueden consultarse en el repositorio de GitHub:
https://github.com/navarro-esm/ULE_data-ai-projects/tree/main

2.2 Evidencias del aprendizaje de los estudiantes (ejemplos, productos, resultados)

Los resultados finales demuestran un alto nivel de dominio en la aplicación de técnicas de ingeniería de datos a la ciencia climática. Los estudiantes pasaron con éxito de scripts básicos a flujos de trabajo complejos y reproducibles que incluyen reducción de dimensionalidad (ACP) y agrupamiento no supervisado (K-Means).

A continuación se incluyen enlaces directos a los proyectos desarrollados por los estudiantes, que constituyen evidencias concretas de los resultados de aprendizaje:

- **Grupo A (Vulnerabilidad agrícola)**: https://github.com/gortif00/CMIP6_Climate_Classification
- **Grupo B (Parque solar):** https://github.com/calejs00/pid
- **Grupo C (Distribución de osos):** https://github.com/SergioAlvesMar5/CMIP6_Climate_Classification_Ivan Fernandez_SergioAlves/

3. Reflexiones y recomendaciones

3.1 Principales logros y fortalezas del proyecto

El principal logro ha sido la integración satisfactoria de la ingeniería de datos 'dura' con la ciencia de la sostenibilidad. Los estudiantes aprendieron a reconocer el carácter transversal del clima: dejaron de ver los conjuntos de datos como meras matrices de números y comprendieron su impacto social y ambiental. Los estudiantes desarrollaron una actitud proactiva, resolviendo los problemas planteados y proponiendo nuevos proyectos que vinculan la ingeniería y el clima, lo que demuestra una comprensión profunda de la materia.

3.2 Dificultades encontradas y cómo se abordaron

Las dificultades encontradas fueron las habituales en una asignatura de nueva implantación, especialmente porque esta es la primera promoción de estudiantes que llega al cuarto curso del grado. Como programa piloto con un grupo de estudiantes completamente nuevo, la asignatura requirió ajustes en los flujos de trabajo técnicos y en el ritmo pedagógico en tiempo real para

garantizar que se alcanzaran los objetivos de aprendizaje sin sobrecargar a los estudiantes. En concreto, se abordaron dos obstáculos técnicos principales:

- **Heterogeneidad de los entornos**: La gestión de entornos de software (Conda) en distintos sistemas operativos (Windows, Mac, Linux) resultó compleja. Herramientas fundamentales como CDO presentaron problemas de compatibilidad en Windows, pero estos se resolvieron orientando a los estudiantes en el uso de subsistemas Linux (WSL) o máquinas virtuales.
- **Estabilidad del acceso a los datos**: Los repositorios externos (Copernicus, ESGF) experimentaron ocasionalmente una latencia elevada o interrupciones del servicio. Para mitigar este problema, se mantuvo la flexibilidad de los plazos y se utilizaron conjuntos de datos de respaldo locales (cachés) para evitar cualquier bloqueo en el progreso del aprendizaje de los estudiantes.

3.3 Lecciones aprendidas para futuras implementaciones

Una de las principales lecciones aprendidas de esta implementación piloto es la importancia de suscitar una implicación inmediata para superar el escepticismo inicial. Aunque los estudiantes mostraron reticencias al principio, su percepción cambió significativamente cuando empezaron a trabajar con datos reales. Por ello, en futuras ediciones se introducirán casos de uso de gran impacto, como fenómenos meteorológicos extremos recientes, desde la primera sesión para captar de inmediato el interés de los estudiantes. Además, la retroalimentación de los estudiantes puso de relieve la necesidad de una mayor flexibilidad metodológica. En futuras implementaciones se dará a los estudiantes más libertad para seleccionar las técnicas estadístico-matemáticas con las que resolver los problemas, fomentando enfoques más creativos y menos guiados una vez dominadas las competencias básicas.

4. Material de apoyo

4.1 Materiales del profesorado y documentación del proyecto

Los materiales del profesorado y la documentación del proyecto están disponibles en el siguiente enlace: https://drive.google.com/drive/folders/17N07p7YOHd7lprLZGN759cnp5EkDBrMi?usp=drive_link

4.2 Entregables elaborados por los estudiantes, presentaciones finales o materiales de presentación breve

Un ejemplo de los entregables elaborados por los estudiantes está disponible en el siguiente enlace: https://drive.google.com/drive/folders/1E7oaA-tSpN8NlM2xiN-W1QScFyMKex4l?usp=drive_link

Las presentaciones finales de los estudiantes están disponibles en el siguiente enlace: https://drive.google.com/drive/folders/1eSmrix7Wk8tS65pVgI1bX6tDfG6RXmAL?usp=drive_link

4.3 Imágenes/vídeos (con consentimiento)

Se realizaron registros fotográficos y de vídeo con el consentimiento explícito de los estudiantes participantes, y están disponibles en el siguiente enlace: https://drive.google.com/drive/folders/1S7-AQBQStS3EH22-l_-x_ZpvM12czDyx?usp=drive_link

4.4 Cualquier otro material pertinente

Todos los materiales, incluido este informe, están disponibles en el siguiente enlace:

https://drive.google.com/drive/folders/1lOJfNdDqlVrzm_YZyGSwbdy6fbT08z4m?usp=drive_link