Curso: Título: Análisis avanzado de datos de experimentación mediante ML e IA

Lugar de impartición:

Modalidad: Dual. Todas las clases se emitirán por streaming para aquellos alumnos que prefieran seguir el curso online. Se facilitará una URL con el enlace a la videoconferencia por correo electrónico unos días antes del comienzo del curso.

Fechas: 8 de mayo, 15 de mayo, 22 de mayo, 29 de mayo, 5 de junio

Duración: 20 horas

Profesorado:

- Pedro Moreno
- Miguel Esteras
- Iker de la Iglesia
- Aitziber Atutxa
- Koldo Gojenola (koldo.gojenola@ehu.eus)

Conocimientos Previos Recomendados:

No se requieren conocimientos previos de programación
Es interesante tener un ordenador accesible para los ejercicios prácticos

**Descripción del Curso:**
Este curso está diseñado para proporcionar a los estudiantes conocimientos fundamentales y habilidades prácticas en el campo del aprendizaje automático (Machine Learning, ML) y la Inteligencia Artificial. Los estudiantes aprenderán a implementar soluciones de ML e IA con tecnologías disponibles y evaluar críticamente estas aplicaciones, midiendo el rendimiento de los modelos y comprendiendo los desafíos que surgen.
El objetivo principal es que los estudiantes desarrollen las habilidades para aplicar las tecnologías de ML e IA a problemas reales

**Metodología:**
Se seguirá un enfoque de aprendizaje basado en problemas en el cual se plantearán problemas prácticos con ejemplos que sirvan como base para resolver problemas de ML e IA.

**Programa del Curso:**

Bloque 1. Flujo de ML (entrenamiento/validación cruzada/interpretabilidad).

Tecnalia (Pedro & Miguel) (16 horas)

- Procesado de datos: Curación, normalización, outliers, PCA/PLS para reducir la dimensionalidad
- Modelos de regresión y clasificación multivariable y optimización bayesiana
- Ajuste y selección de modelos, Interpretabilidad (SHAP)
- Validación de modelos e intervalos de confianza

Bloque 2. Grandes Modelos de Lenguaje (LLMs). UPV/EHU (4 horas)

- Introducción a los LLMs
- Zero-shot, One-shot, Few-shot, Chain of Thought, RAG
- Agentes
- Aplicaciones en el dominio de la bioinformática y la salud

## Calendario del Curso:

DIA 1: Procesado de datos: Curación, normalización, outliers, sesgos, y redución de dimensionalidad – 4 HORAS
DIA 2: Modelos de regresión y clasificación, y aprendizaje no supervisado – 4 HORAS
DIA 3: Ajuste y selección de modelos, y interpretabilidad – 4 HORAS
DIA 4: Validación de modelos e intervalos de confianza – 4 HORAS
DIA 5 - LLMs– 4 HORAS

## Material online

El curso incluye material online (documentación).

SISTEMA DE EVALUACIÓN
1. Evaluación Continua (100% de la calificación)
Se basará en el trabajo y ejercicios prácticos realizados en clase