

# Técnicas de aprendizaje automático y minería de datos: Nivel iniciación

**Eje estratégico:** Formación en competencias para la investigación.

**Modalidad de formación:** Presencial

**Colectivo al que se dirige:** Personal docente e investigador

**Plazas ofertadas:** 15

**Requisitos previos:** Se recomienda tener unos conocimientos básicos de programación, sin ningún lenguaje de programación en particular. Conocimientos de estadística. Asimismo, es conveniente conocer los indicadores estadísticos básicos (media, desviación típica, mediana, cuantiles, etc.), concepto de muestreo y nociones básicas de estimación de modelos (Modelo de regresión lineal).

**Duración estimada de la formación:** 10 horas

**Calendario y planificación:** 10 y 11 de enero de 2024, de 15:00 a 20:00 horas.

**Persona formadora:** D<sup>a</sup> Yolanda Rebollo Sanz, Profesora Titular de Universidad en el Departamento de Economía, Métodos Cuantitativos e Historia Económica.

## Objetivos de la acción formativa:

Las técnicas de minería de datos (data mining) y aprendizaje automático (Machine Learning) tienen como objetivo esencial el descubrimiento automático del conocimiento contenido en la información almacenada de modo ordenado en conjuntos de datos que pueden tener una dimensión elevada. Estas técnicas tienen por objeto descubrir patrones, perfiles, tendencias y otras relaciones presentes en la información, pero ocultas si no se tratan adecuadamente.

La temática de este curso se centra en las técnicas de minería de datos y aprendizaje automático y su implementación mediante software específico (STATA) que se puede usar en diferentes áreas de conocimiento.

El curso está destinado a las personas interesadas en la aplicación de técnicas avanzadas de análisis y procesamiento de datos, como el Aprendizaje Automático, y que quieran desarrollar habilidades prácticas para aplicarlas en el mundo real. Es un curso de iniciación diseñado para ser accesible a principiantes en esta área temática y que les puede ayudar a mejorar sus habilidades para trabajar con datos.

## **Programa de contenidos:**

### **1. Introducción: Conceptos fundamentales en el aprendizaje automático y minería de datos**

Presentación de los conceptos fundamentales (Funciones de pérdida, Sobreajuste, Matriz de confusión, el problema de la dimensión, Validación Cruzada etc) que subyacen en las técnicas de aprendizaje automático.

### **2. Iniciación al Aprendizaje automático supervisado (predicción y clasificación)**

Presentación de las técnicas principales relacionadas con el aprendizaje automático supervisado aplicado al objetivo de predicción (Técnicas y Aplicaciones con Software STATA).

**2.1** Aprendizaje Automático Supervisado: Parametric Regresion

**2.2** Aprendizaje Automático Supervisado: Clasificación

En particular se analizan los tipos de problemas distinguiendo entre la clasificación y la predicción de valores y se analizan las técnicas básicas o algoritmos de Penalización (Regularized Regression), árboles de decisión y bosques aleatorios.

Entre los algoritmos de penalización se encuentran las técnicas de Lasso, Elastic Net y Ridge Regresión.

En todos los casos, se analizará también como evaluar los resultados obtenidos y como comparar los resultados obtenidos usando distintas técnicas. Se aprenderán las técnicas que se deben aplicar y por qué.

### **3. Iniciación al Aprendizaje automático no supervisado (predicción y clasificación)**

**3.1** Técnicas de Agrupamiento (Clustering)

Presentación de las técnicas principales relacionadas con el aprendizaje automático no supervisado (Técnicas y Aplicaciones con Software STATA).

El objetivo general de esta sección del curso es familiarizarse con las técnicas básicas de aprendizaje No supervisado. En particular se analizan los tipos de problemas distinguiendo entre los distintos tipos de clustering y el problema de calcular la similitud y se analizan las técnicas básicas para obtener clusterings. Se analizará también como evaluar los resultados obtenidos.

Aunque proporcionaremos una breve introducción a STATA, el enfoque de la clase estará en enseñar conceptos asociados a las técnicas de aprendizaje automático en lugar de como usar STATA.

## **Metodología:**

El curso tiene como objetivo principal el de potenciar el aprendizaje, facilitando la adquisición de los conocimientos y competencias contempladas en el programa. Dadas las características del curso, los métodos a utilizar variarán: expositivo, por descubrimiento, interactivo, enseñanza colaborativa, enseñanza programada, trabajo autónomo, metodología participativa (trabajos grupales, aprendizaje cooperativo, ...), etc.

Al comenzar cada módulo del curso, el o la estudiante dispondrá del material correspondiente. Se incluirán las diapositivas presentadas en clase, como guion de los conocimientos a adquirir, para facilitar su seguimiento. Se proporcionarán ejemplos empíricos que pueden servir también como hojas de ejercicios, así como toda la documentación que se considere necesaria relativas a los contenidos formativos. Los diferentes métodos docentes que se emplearán a lo largo del curso son:

1. Sesiones magistrales donde se expone la teoría básica del programa en cada uno de los temas que se vayan a tocar.
2. Sesiones prácticas donde se realizará un trabajo práctico propuestos por la profesora y relacionados con la materia estudiada.

Los contenidos se desarrollarán en el aula (clase magistral + prácticas en aula), resolviendo supuestos teóricos prácticos. Cada sesión constará de explicaciones teóricas seguidas de actividades prácticas. La explicación teórica se estructurará, en dos partes de manera que se pueda practicar inmediatamente lo explicado. Durante la parte teórica se introducirán los conceptos clave y se presentarán los elementos tecnológicos más relevantes de cada tema, y durante la parte práctica se aplicarán estos conceptos a través de una serie de ejercicios con bases de datos propuestos.

### **Resultados de aprendizaje:**

En este curso se aprenden los conceptos básicos del aprendizaje automático (machine learning); en particular, las técnicas necesarias para evaluar el rendimiento de los algoritmos y de los modelos obtenidos. También se aprenderá cómo preprocesar los datos para obtener así modelos de mayor calidad (simples, comprensibles, eficientes, etc.), cómo utilizar técnicas de aprendizaje automático, para extraer modelos descriptivos y predictivos a partir de los datos, así como saber evaluarlos correctamente, y cómo utilizar un lenguaje de programación de análisis de datos (STATA) que permiten generar modelos de aprendizaje automático.

**Forma de participación:** Se fomentará el aprendizaje participativo del alumnado a lo largo del curso. La parte teórica ha sido diseñada con numerosos ejemplos que buscan motivar a reflexionar y proponer enfoques de análisis. En la parte práctica del curso, las personas participantes experimentarán individualmente con las técnicas de análisis aprendidas en la sección teórica. Esto les permitirá aplicar los conocimientos adquiridos y fortalecer su comprensión práctica de las técnicas de aprendizaje automático.

**Tipo de evaluación:** Para la evaluación del alumno se sigue un tipo criterial, i.e. una evaluación que pretende determinar las competencias que posee el estudiantado tras el desarrollo del proceso de aprendizaje. En este tipo de evaluación se recoge información mediante un instrumento, prueba, procedimiento o actividad para poder describir e identificar las competencias adquiridas por los estudiantes acerca de un dominio de referencia, descrito en la planificación docente inicial.

**Criterios para acreditar la formación realizada:** Superación de la evaluación.

**Persona formadora:** D<sup>a</sup> Yolanda Rebollo Sanz, Profesora Titular de Universidad en el Departamento de Economía, Métodos Cuantitativos e Historia Económica.

Doctora en Economía por la Universidad Pablo de Olavide, Máster en Análisis Económico por la Universidad Pompeu Fabra y el Instituto de Estudios Fiscales. Es miembro e investigador principal en diferentes grupos de investigación especializados en métodos cuantitativos y evaluación de políticas públicas. Actualmente es Profesora Titular en el Departamento de Economía de la Universidad Pablo de Olavide de Sevilla. Es profesora titular en el área de Análisis Económico del Departamento de Economía, Métodos Cuantitativos e Historia Económica de la Universidad Pablo de Olavide. Ha desarrollado su actividad docente e investigadora en dicha universidad desde el año 1998. Cuenta con una dilatada experiencia docente e investigadora. Es autora de un amplio número de investigación con factor de Impacto (JCR, SJR) donde se destaca el

uso de métodos cuantitativos avanzados (Microeconometría Avanzada e Inferencia Causal) para la mejor comprensión de la realidad económica.