

Big Data. Análisis de datos y aprendizaje automático con Python

Cód. B01

DIRECCIÓN:

Guadalupe Miñana Ropero y Yolanda García Ruiz.

ÁREA EN LA QUE SE INSCRIBE EL CURSO:

Ciencias Experimentales.

HORARIO DEL CURSO:

Mañanas de 9:00 a 14:00 horas, de lunes a viernes.

NÚMERO DE ALUMNOS:

20.

PERFIL DEL ALUMNO:

Estudiantes universitarios, docentes, investigadores y profesionales de cualquier sector con interés en la temática del curso. No se requieren conocimientos previos.

OBJETIVOS:

- Conseguir una visión global del amplio ecosistema Big Data y la inteligencia artificial, así como descubrir los retos y las oportunidades de sus aplicaciones.
- Entender la nueva manera de obtener, gestionar y analizar datos para generar conocimiento e incorporarlo a la toma de decisiones.
- Poder reflexionar, analizar y abrir la mente al cambio de sociedad que estamos viviendo para entender su impacto en la sociedad actual y encontrar nuevas profesiones, retos y oportunidades.
- Presentar el lenguaje Python, ampliamente utilizado en el contexto del análisis de datos debido a que dispone de un gran número de librerías que incluyen herramientas para cubrir el ciclo completo del dato: captura de datos, almacenamiento, preprocesamiento y limpieza, visualización y análisis de resultados.
- Mostrar distintas técnicas de aprendizaje automático para el diseño de modelos que permitan obtener información valiosa a partir de un conjunto de datos para ayudar a la toma de decisiones. Creación de modelos de aprendizaje con Python. Interpretación y validación de los modelos creados. Valorar qué modelo es el más apropiado y ajustarlo.

PROGRAMA:

- **Introducción al Big Data.**

- Introducción al ecosistema Big Data.
- Casos de uso del Big Data.
- Inteligencia artificial y su relación con Big Data.
- Open data.
- Ciclo de un proyecto Big Data.
 - Fuentes de datos, adquisición, limpieza, transformación e integración.
 - Almacenamiento: bases de datos NoSQL.
 - Análisis y modelado. Ingeniería de características. Aprendizaje automático: modelos de regresión, árboles de decisión, random forest, redes neuronales.
 - Visualización.
- **Análisis de datos con Python.**
 - Introducción al lenguaje de programación Python.
 - Carga y almacenamiento de datos.
 - Procesamiento y limpieza de datos con Pandas.
 - Visualización de datos con Matplotlib y Plotly.
- **Técnicas de Aprendizaje automático con Python.**
 - Introducción al aprendizaje automático.
 - Modelos de aprendizaje supervisado. Regresión y clasificación.
 - Optimización de hiperparámetros. Validación de modelos.
 - Modelos de aprendizaje no supervisado.

ACTIVIDADES PRÁCTICAS:

- Se realizarán diferentes proyectos guiados usando el lenguaje Python para cubrir el ciclo de vida del dato: desde la captura de datos, hasta la creación de modelos de aprendizaje.

PROFESORADO:

- D^a Guadalupe Miñana Roperó, UCM.
- Yolanda García Ruiz, UCM.
- Guillermo Jiménez Díaz, UCM.