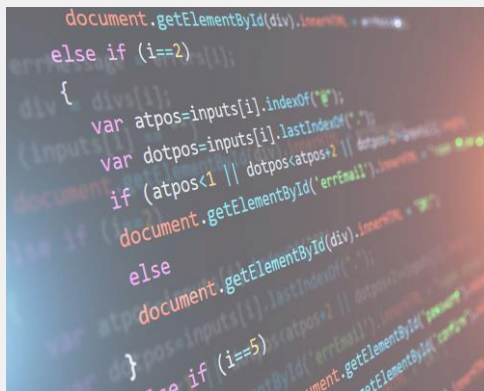




Presentación curso Ciencia de Datos

01



Presentación

Presentación del curso Ciencia de Datos.

02

Contenidos del curso

Funcionamiento del curso y contenido de los distintos módulos que lo componen.

03

Sistema de aprendizaje

Sistema de aprendizaje y entrega de las tareas y calendario propuesto.

Presentación del curso

El objetivo principal de este curso es conocer los conceptos básicos de ciencia de datos y conseguir una competencia efectiva con código Python, orientado a la analítica de datos.

El curso está orientado a alumnos con conocimientos básicos de estadística, SQL y Python.

Se trabaja en el entorno Jupyter Notebook. En la primera parte del curso se trabajarán las bases de datos con MySQL, las herramientas de programación numérica, los dataframes (similares a tablas de excel) y los paquetes de visualización de datos.

En los siguientes módulos se introducirá Machine Learning y se trabajarán los algoritmos de aprendizaje supervisado y no supervisado. Por último se presentarán herramientas como los pipelines, herramientas de texto minning y técnicas de web scraping.

Una vez finalizados los diferentes módulos y presentadas las tareas asociadas, los alumnos pasarán a realizar un proyecto final transversal en el que aplicar los conocimientos adquiridos.

Contenidos del curso

El curso cuenta con un itinerario de 10 módulos temáticos y un proyecto final transversal. Cada uno de estos módulos cuenta con distintos recursos de autoaprendizaje y tareas asociadas a cada uno de los módulos.

- Módulo 0. Presentación del curso e instalación del entorno de trabajo : en este módulo se explicará qué es y para que sirve un entorno de trabajo y se introducirá e instalará el entorno Jupyter Notebook y Github para realizar la entrega de los ejercicios.
- Módulo 1. Bases de datos relacionales (SQL) : en este módulo se verán conceptos básicos de SQL y se crearán y consultarán bases de datos relacionales con MySQL.
- Módulo 2. Básicos de Python: se trabajan conceptos básicos de Python como las estructuras de datos y las estructuras de control.
- Módulo 3. Programación numérica, dataframes y análisis estadístico: en este módulo se trabajan los paquetes básicos de Python que permiten un tratamiento estadístico de los datos así como el trabajo y disposición en tablas. Se presentan las librerías NumPy y Pandas.

Contenidos del curso

- Módulo 4. Visualización gráfica de datos : en este módulo se trabajan los principales paquetes que permiten visualizar los datos gráficamente. En concreto, los paquetes Matplotlib y Seaborn.
- Módulo 5. Introducción al Machine Learning : se introducen los primeros conceptos de machine learning y se presenta uno de los paquetes principales del data science con python: scikitlearn.
- Módulo 6. Algoritmos de aprendizaje supervisado: Regresión: en este módulo se presentan y trabajan los principales algoritmos de regresión, su presentación y evaluación.
- Módulo 7. Algoritmos de aprendizaje supervisado: Clasificación: en este módulo se presentan y trabajan los principales algoritmos de clasificación, su presentación y cómo se evalúan.
- Módulo 8. Algoritmos de aprendizaje NO supervisado: Agrupación : se presentan y trabajan los principales algoritmos de agrupación, cómo se presentan y cómo se evalúan.
- Módulo 9. Análisis de sentimiento y textos : en este módulo se introduce y se trabaja el texto minning.
- Módulo 10. Web scraping y automatización: en este módulo se presentan y trabajan los principales paquetes de web scraping por Python y se pone en práctica con diferentes ejercicios.

Contenidos del curso

Proyecto final : el proyecto final debe tener una vertiente profesionalizadora. Esto significa que debe estar enfocado a una casuística o interés real. El proyecto debe contar, como mínimo, con los siguientes apartados:

- Presentación, contextualización e interés del caso.
- Descripción de los datos.
- Análisis estadístico y gráfico de los datos más relevantes.
- Aplicación de algoritmos vistos a lo largo del curso y que puedan resultar de utilidad al caso.
- Análisis de la información extraída.
- Conclusiones

Se recomienda empezar a trabajar la idea del proyecto en torno a los módulos 5 y 6 para ir definiendo con el mentor/a el alcance del mismo.

Este proyecto deberá presentarse y defenderse en un encuentro virtual con el mentor/a.

Sistema de aprendizaje

La metodología del curso se basa en 4 pilares principales:

- **Autoaprendizaje:** visualizar los distintos recursos disponibles (vídeos, tutoriales, recursos web, etc.) y realizar diferentes ejercicios y pruebas para consolidar el aprendizaje.
- **A tu modo:** un sistema de aprendizaje versátil. Se pueden visualizar los recursos disponibles, si son necesarios, para realizar los ejercicios propuestos en el momento que se desee.
- **Práctico:** no se estudian conceptos ya conocidos por el alumno sino que se introducen y se demuestran aplicándolos en ejercicios consolidando e incorporando nuevos conocimientos.
- **Mentoría:** todas estas tareas se realizan de forma autónoma, pero con un mentor/a con total disposición para ayudar a resolver todo aquello que no pueda hacer el estudiante. Dentro del moodle existe un foro para plantear todas las dudas que vayan surgiendo tanto en abierto como directamente al mentor/a.

El mentor/a también está disponible en el aula virtual, por correo electrónico y en el canal de Discord.

Sistema de aprendizaje

Las **tareas** son el centro del curso. Hay que focalizar todos los esfuerzos en realizar las tareas de la mejor manera posible.

La entrega de las tareas se realiza a través del moodle. En algunas tareas tendrá que entregar el archivo directamente y en otras tendrá que entregar un archivo de texto que contenga el enlace a su repositorio de Github donde debe colgar los ejercicios para que se corrijan.

Dentro de cada tarea, en cada módulo, se especifica el sistema de entrega a realizar.

Una vez realizada la entrega, el mentor/a revisará sus ejercicios y se los corregirá reportándole un feedback sobre cómo se han hecho los ejercicios.

Es importante recordar que el mentor/a tiene diferentes vías de contacto y que es un recurso que puede ser usado en cualquier momento por cualquier alumno que tenga una duda o quiera realizar una consulta.