MASTER DATA SCIENCE ONLINE. PYTHON, SQL & IA

Duración: 250 horas (5 meses)

Modalidad: Online Teleformación

Público: Desarrolladores y analistas de datos que deseen aprender a desarrollar en Python proyectos de Data Science, Machine Learning y Deep Learning a nivel empresarial.

Objetivos:

- Aprender a programar en Python y conocer sus principales librerías.
- Adquirir los conocimientos de la programación orientada a objetos
- Comprender y profundizar en el flujo completo de un proyecto de Data Science para convertirse en científico de datos
- Aprender a realizar consultas en bases de datos SQL
- Dominar el Machine Learning y qué modelo y algoritmo utilizar para cada reto, siendo capaz de resolver problemas de clasificación, regresión, clustering y reglas de asociación.
- Aprender los conceptos fundamentales del Deep Learning y conocer los diferentes tipos de Redes Neuronales, valorando cuál es el más adecuado y optimizarlos.
- Aprender a utilizar técnicas avanzadas de Procesamiento de Lenguaje Natural (NLP) implementadas en Python.
- Aprender a implementar modelos de Machine Learning y Deep Learning para la clasificación y predicción de tendencias futuras, fomentando una mejor toma de decisiones en la empresa y una mayor competitividad en el mercado.

TEMARIO

MÓDULO 1. Python Introducción (40 h)

- ¿Por qué aprender Python?
  - El lenguaje Python en el mundo actual
  - Aplicaciones de Python en el desarrollo de software
  - Características generales del lenguaje
- Primeros pasos con Python
  - Herramientas y entornos de desarrollo
  - Configurando el entorno de desarrollo
  - Características de Python
  - Primer programa Python

- Sintaxis de Python
  - Tipos de datos y variables
  - Entrada y salida
  - Instrucciones de control de flujo
  - Principales funciones del lenguaje
  - Pip y utilización de módulos Python
  - Manejo de cadenas y fechas
- Estructuras de datos
  - Listas y tuplas
  - Diccionarios
  - Utilización de lambdas
  - Manejo de cadenas y fechas
- Programación orientada a objetos con Python
  - Clases y objetos
  - Atributos
  - Constructores
  - Herencia
  - Sobreescritura de métodos
  - Metaclases y decoradores
  - Uso de type
  - Excepciones
- Acceso a ficheros y bases de datos
  - Lectura y escritura en ficheros
  - Módulos para acceso a ficheros
  - Acceso a ficheros CSV
  - Acceso a ficheros JSON
  - Bases de datos relacionales y documentales
  - Módulo para acceso a bases de datos MySQL
  - Acceso a bases de datos MongoDB
- Librerías para manipulación de datos
  - Principales librerías para el tratamiento de datos en Python
  - Utilización de la librería Numpy
  - Manipulación de arrays con Numpy
  - Pandas

MÓDULO 2. Python Avanzado (40 h)

- Funciones avanzadas
  - Listas
  - Ejercicios de listas
  - Comprensión de listas
  - Diccionarios
  - Funcionarios Lambdas

- Programación orientada a objetos
  - Clases y objetos
  - Destructores y atributos
  - Herencia
  - Ejercicio de herencia
  - Sobreescritura
  - Métodos especiales
- Multitarea
  - Fundamentos de multitarea
  - Multitarea con threading
  - Sincronización de tareas
- Metaprogramación
  - Conceptos generales de metaprogramación
  - Manipulación de clases
  - Metaclases
  - Creación de metaclases
  - Decoradores
- Módulos y paquetes
  - Módulos I
  - Módulos II
  - Paquetes
  - Distribución de paquetes
- Acceso a datos
  - Ficheros
  - Ficheros y otros métodos
  - Ejercicio de gestión de empleados
  - Manipulación JSON
  - Gestión de empleados JSON
  - Fundamentos MongoDB
  - Inserción de documentos MongoDB
  - Recuperación, actualización, eliminación MongoDB
  - Ejercicio MongoDB
- Acceso a APIs
  - Acceso a recursos externos
  - Manipulación de respuesta
  - API REST
  - Ejemplos APIS
  - Autentificación APIS
- Scripts de administración
  - Módulos SYS
  - Módulo Subprocess
  - Salida subprocesos

- El módulo OS

MÓDULO 3. Python & SQL (20 h)

- Lenguaje SQL y gestores de bases de datos en Python
- Profundizando en SQLite: El gestor SQL más rápido
  - Introducción a SQLite
  - Instalación y primeras conexiones
  - Creación, edición y eliminación de bases de datos y tablas
  - Operaciones esenciales de manipulación de datos (CRUD: Create, Read, Update, Delete)
  - Consultas SQL esenciales (selects)
  - Relaciones entre tablas
  - Consultas SQL avanzadas (selects y joins)
  - Importación y exportación
- Conexión de SQLite a Power BI usando ODBC

MÓDULO 4. Python & Data Science (45 h)

- Introducción al análisis de datos
  - ¿Qué es Python y qué nos proporciona para el análisis de datos?
  - Instalación Python + Jupyter
  - Importar librerías y fuentes de datos
  - Visualización básica con Matplotlib
  - Flujograma de un proyecto Data Science
- Fundamentos del lenguaje Python
  - Variables en Python
  - Creación de listas y extracción de datos
  - Conceptos avanzados de creación de listas
  - Uso de funciones en Python (in-built)
  - Creación de funciones en Python y argumentos flexibles
  - Funciones lambda
  - Métodos de Python
  - Cómo crear diccionarios en Python
  - Uso de función zip para creación de diccionarios en base a listas
  - Operadores en Python
  - Bucles en Python
  - Comprensión de listas en Python
- Conceptos de estadística para el análisis de datos
  - Variables y conceptos básicos
  - Varianza en una variable
  - Correlación de variables
  - Histogramas
  - Análisis con percentiles (CDF)

    - Funciones densidad de probabilidad
    - Cálculo de previsiones (forecast) y media móvil
- Cálculo numérico con Numpy
    - Introducción a la librería Numpy
    - Selección de datos con array Numpy
    - Arrays 2D en Numpy
    - Cálculo estadístico con Numpy
- Análisis de datos con Pandas
    - Introducción a la librería Pandas, ¿qué es un dataframe?
    - Creación de un dataframe a partir de un diccionario
    - Cómo importar datos desde un fichero de texto plano (txt, csv...)
    - Selección de datos en un dataframe Pandas
    - Métodos útiles de un dataframe Pandas
    - Eliminar duplicados, valores erróneos y columnas de un dataframe Pandas
    - Interpolación de datos
    - Filtrar datos en un dataframe Pandas
    - Ordenación de valores en un dataframe Pandas
    - Crear columnas en un dataframe para cadenas de texto
    - Crear columnas en un dataframe a partir de un diccionario con map
    - Crear columnas en un dataframe a partir de funciones lambda
    - Crear columnas en un dataframe a partir de funciones condicionales
    - Renombrar y reordenar columnas de un dataframe Pandas
    - Cómo crear pivot tables en Pandas
    - Uso de groupby en Pandas
    - Concatenación de dataframes (union)
    - Combinación de dataframes
- Importación y exportación con Pandas
    - Cómo importar datos desde un fichero Excel
    - Introducción a las BBDD relacionales / Modelos de datos
    - Cómo importar datos desde una BBDD SQL
    - Cómo importar datos desde una página web
    - Cómo importar datos desde una página web (web scraping)
    - Cómo importar datos desde un fichero semi-estructurado JSON
    - Cómo importar datos desde redes sociales
    - Cómo importar datos desde cloud (AWS / Azure / Google Cloud)
    - Exportación de datos a csv y Excel
    - Exportación de datos a BBDD SQL
- Visualización de datos en Python. Matplotlib
    - Consejos de visualización de datos
    - Introducción a la librería Matplotlib
    - Creación de un gráfico de línea, bar, scatter

  - Personalización de gráficos (título, etiquetas, ticks, leyenda, límites de ejes, anotaciones…)
  - Creación de box & whiskers plot
  - Creación de un histograma y CDF
  - Gráfico de media móvil
  - Visualización de gráficos múltiples (subplots)
  - Aplicación de estilos
  - Creación de gráficos a partir de un objeto groupby
  - Creación de histogramas en 2D
  - Creación de mapas geográficos con basemap y Google maps
- Visualización de datos en Python. Seaborn
  - Introducción a la librería Seaborn
  - Seaborn – Creación de regresión lineal
  - Seaborn – Stripplot
  - Seaborn – Swarmplot
  - Seaborn – Violinplot
  - Seaborn – Uso de jointplot
  - Seaborn – Uso de pairplot
  - Seaborn – Correlación con heatmap
- Series temporales en Python
  - Series temporales en Pandas: extracción y parsing
  - Series temporales: filtrado
  - Series temporales: remuestreo – diezmado (downsampling)
  - Series temporales: remuestreo – interpolación (upsampling)
  - Visualización de series temporales
  - Previsiones basadas en datos históricos
- Ejecución e interconexión de datos de Python con otras plataformas
  - Generación de scripts de Python y automatización de tareas
  - Uso de Python en herramienta de business intelligence Power BI

MÓDULO 5. Python & Machine Learning (25 h)

- Introducción a Machine Learning
  - ¿Qué es el Machine Learning y la Ciencia de Datos?
  - Machine Learning, Inteligencia Artificial y Deep Learning
  - Tipos de Machine Learning
- Introducción a Python
  - Instalación Python + Jupyter
  - Conceptos básicos de Python
  - Introducción a las librerías: Numpy
  - Introducción a las librerías: Pandas
  - Introducción a las librerías: Matplotlib
  - Librería Machine Learning Scikit-Learn

- Machine Learning. Clasificación
  - ¿Qué es y qué problemas resuelven un modelo de clasificación?
  - Algoritmos Machine Learning para Clasificación (Decision Tree, SVM, Naive Bayes, Logistic Regression. KNN)
  - Explicación paso a paso con Scikit-Learn. Caso práctico de Clasificación
- Machine Learning. Regresión
  - ¿Qué es y qué problemas resuelven un modelo de Regresión?
  - Algoritmo Machine Learning Regresión Lineal
  - Explicación paso a paso con Scikit-Learn. Caso práctico de Regresión
- Machine Learning. Clustering
  - ¿Qué es y qué problemas resuelven un modelo de Clustering?
  - Algoritmo Machine Learning K-Means
  - Explicación paso a paso con Scikit-Learn. Caso práctico de Clustering
- Reglas de asociación
  - ¿Qué es y qué problemas resuelven un modelo de Reglas de Asociación?
  - Algoritmo Reglas de Asociación - «Apriori»
  - Explicación paso a paso. Caso práctico de Reglas de Asociación.

MÓDULO 6. Python & Deep Learning (40 h)

- Introducción a Deep Learning
  - ¿Qué es Machine Learning y Deep Learning?
  - Instalación de entorno Python y librerías Deep Learning
  - Aprendizaje supervisado
  - ¿Qué es el overfitting / underfitting en el aprendizaje supervisado?
  - Evaluación de rendimiento de modelos - Métricas de Clasificación
  - Evaluación de rendimiento de modelos - Métricas de Regresión
  - Aprendizaje no supervisado
- Redes neuronales artificiales (ANN). Conceptos fundamentales
  - ¿Qué es una neurona y el modelo perceptrón?
  - ¿Qué son las redes neuronales?
  - Funciones de activación
  - Funciones de activación en modelos multiclase
  - Funciones de Coste y de Gradiente Descendente
  - Propagación hacia atrás (backpropagation)
  - Claves para crear redes neuronales efectivas
  - ¿Qué nos proporciona Tensorflow y Keras?
- Redes neuronales artificiales (ANN). Regresión con Keras y Tensorflow
  - Regresión con Keras - Presentación caso práctico
  - Regresión con Keras - Importación de librerías y fuentes
  - Regresión con Keras - Análisis de datos (EDA) + Preprocesado (I)
  - Regresión con Keras - Análisis de datos (EDA) + Preprocesado (II)
  - Regresión con Keras - División Train / Test

    - Regresión con Keras - Escalado
    - Regresión con Keras - Creación de modelo
    - Regresión con Keras - Entrenamiento del modelo
    - Regresión con Keras - Evaluación y Predicción
- Redes neuronales artificiales (ANN). Clasificación binaria con Keras y Tensorflow
    - Clasificación binaria con Keras - Presentación caso práctico
    - Clasificación binaria con Keras - Importación de librerías y fuentes
    - Clasificación binaria con Keras - Análisis de datos (EDA) + Preprocesado
    - Clasificación binaria con Keras - División Train / Test
    - Clasificación binaria con Keras - Escalado
    - Clasificación binaria con Keras - Creación de modelo
    - Clasificación binaria con Keras - Entrenamiento del modelo
    - Clasificación binaria con Keras - Evaluación y Predicción
- Redes neuronales artificiales (ANN). Clasificación multiclase con Keras y Tensorflow
    - Clasificación multiclase con Keras - Presentación caso práctico
    - Clasificación multiclase con Keras - Importación de librerías y fuentes
    - Clasificación multiclase con Keras - Análisis de datos (EDA) + Preprocesado (I)
    - Clasificación multiclase con Keras - Análisis de datos (EDA) + Preprocesado (II)
    - Clasificación multiclase con Keras - División Train / Test
    - Clasificación multiclase con Keras - Escalado
    - Clasificación multiclase con Keras - Creación de modelo
    - Clasificación multiclase con Keras - Entrenamiento del modelo
    - Clasificación multiclase con Keras - Evaluación y Predicción
    - Clasificación multiclase con Keras - Monitorización con Tensorboard
- Redes neuronales convolucionales (CNN). Clasificaciones imágenes blanco y negro
    - Introducción a las redes neuronales convolucionales (CNN)
    - ¿Qué son los filtros de imagen y los kernels?
    - Capas convolucionales en una CNN
    - Capas pooling en una CNN
    - Clasificación imágenes Blanco y Negro - Presentación caso práctico
    - Clasificación imágenes Blanco y Negro - Importación de librerías y fuentes
    - Clasificación imágenes Blanco y Negro - Preprocesado
    - Clasificación imágenes Blanco y Negro - Creación del modelo
    - Clasificación imágenes Blanco y Negro - Entrenamiento del modelo
    - Clasificación imágenes Blanco y Negro - Evaluación y Predicción
- Redes neuronales convolucionales (CNN). Clasificaciones imágenes en color
    - Clasificación imágenes RGB - Presentación caso práctico
    - Clasificación imágenes RGB - Importación de librerías y fuentes
    - Clasificación imágenes RGB - Preprocesado
    - Clasificación imágenes RGB - Creación del modelo
    - Clasificación imágenes RGB - Entrenamiento del modelo
    - Clasificación imágenes RGB - Evaluación y Predicción

- Redes neuronales recurrentes (RNN)
  - Introducción a las redes neuronales recurrentes (RNN)
  - Neuronas LSTM
  - Creación de batches en RNN
  - Forecast RNN - Presentación caso práctico
  - Forecast RNN - Importación de librerías y fuentes
  - Forecast RNN - Preprocesado
  - Forecast RNN - División Train / Test
  - Forecast RNN - Escalado
  - Forecast RNN - Creación Generador Serie Temporal
  - Forecast RNN - Creación del modelo
  - Forecast RNN - Entrenamiento del modelo
  - Forecast RNN - Evaluación y Predicción
- Redes neuronales en Aprendizaje No Supervisado
  - Introducción a las redes neuronales en aprendizaje no supervisado
  - ¿Qué son los autoencoders en una red neuronal?
  - NN No Supervisado - Presentación caso práctico
  - NN No Supervisado - Importación de librerías y fuentes
  - NN No Supervisado - Preprocesado
  - NN No Supervisado - Escalado
  - NN No Supervisado - Estimación número de clústeres
  - NN No Supervisado - Creación del modelo
  - NN No Supervisado - Entrenamiento del modelo
  - NN No Supervisado - Evaluación y Predicción de clústeres

MÓDULO 7. NLP (Procesamiento de Lenguaje Natural) en Python (10 h)

- ¿Qué es el Procesamiento del Lenguaje Natural?
- Librerías en Python de NLP: NLTK, Gensim, Huggin Face y Spacy
- Corpus del lenguaje en Spacy
- Tokenización y análisis léxico
  - Tokenización: Dividir texto en palabras y frases
  - Reconocimiento de entidades nombradas (NER): Qué son y cómo identificarlas
  - Part-of-Speech (POS): Clasificando palabras por su función gramatical
- Procesamiento avanzado del texto
  - Lemmatización: Reduciendo palabras a su forma base
  - Dependencias sintácticas: Relaciones entre palabras y contexto
  - Embeddings de palabras: Representando texto en vectores multidimensionales
- Aplicaciones prácticas de NLP: Análisis de sentimientos, extracción de información o generación de texto

MÓDULO 8. Trabajo Fin de Máster (30 h)

- Desde el análisis hasta la predicción: Aplicaciones de Python, SQL y técnicas de IA (ML, DL, NLP).
    - El Trabajo Fin de Máster se centrará en desarrollar un sistema que recopile y analice las reseñas y opiniones de los clientes de la empresa, utilizando técnicas avanzadas de Procesamiento de Lenguaje Natural (NLP) implementadas en Python. El objetivo principal es extraer información valiosa que permita identificar áreas de mejora y adaptar los servicios o productos a las necesidades y expectativas de los consumidores. Para ello, se emplearán herramientas como spaCy para el preprocesamiento y análisis de los textos, y pandas para la manipulación y gestión eficiente de los datos. Durante el proyecto, se integrarán bases de datos relacionales mediante SQL para almacenar y gestionar eficientemente las opiniones recopiladas, facilitando su posterior análisis.
    - Además, se implementarán modelos de Machine Learning y/o Deep Learning para clasificar y predecir tendencias en las opiniones de los consumidores, permitiendo a la empresa tomar decisiones informadas para mejorar la calidad de sus servicios o productos. Este enfoque contribuirá a optimizar la satisfacción del cliente y, en última instancia, a incrementar la competitividad en el mercado.