

# Análisis de datos con Python

## Modelización y representación de datos

Curso práctico de 4 días - 28h

Ref.: BDA - Precio 2024: 2 000€ sin IVA

Data Analytics es un término utilizado para expresar el proceso de análisis de datos con el fin de tomar decisiones. El lenguaje Python ofrece un ecosistema para el tratamiento estadístico, desde la construcción de modelos de análisis hasta su evaluación y representación.

### OBJETIVOS PEDAGÓGICOS

Al término de la formación, el alumno podrá:

Comprender el principio de la modelización estadística

Elegir entre regresión y clasificación en función del tipo de datos

Evaluación del rendimiento predictivo de un algoritmo

Crear selecciones y clasificaciones a partir de grandes volúmenes de datos para identificar tendencias

Desarrollo/creación de análisis utilizando software Python, con los módulos pandas, NumPy, SciPy, Matplotlib, seaborn, scikit-learn y statsmodels.

## PROGRAMA

última actualización: 09/2023

### 1) Introducción a la modelización

- Introducción al lenguaje Python.
- Introducción al software Jupiter Notebook.
- Las etapas de la construcción de un modelo.
- Algoritmos supervisados y no supervisados.
- La elección entre regresión y clasificación.

*Trabajo práctico* : Instalación de Python 3, Anaconda y Jupiter Notebook.

### 2) Procedimientos de evaluación de modelos

- Técnicas de remuestreo en juegos de entrenamiento, validación y prueba.
- Comprobación de la representatividad de los datos de formación.
- Medición del rendimiento de los modelos predictivos.
- Matriz de confusión y de costes, curvas ROC y AUC.

*Trabajo práctico* : Establecer el muestreo de conjuntos de datos. Realizar pruebas de evaluación de varios modelos suministrados.

### 3) Algoritmos supervisados

- El principio de la regresión lineal univariante.
- Regresión multivariante.
- Regresión polinómica.
- Regresión regularizada.
- El Bayes ingenuo.
- Regresión logística.

*Trabajo práctico* : Aplicación de regresiones y clasificaciones en varios tipos de datos.

### PARTICIPANTES

Gestores de centros de datos (minería de datos, marketing, calidad, etc.), usuarios de bases de datos y gestores empresariales.

### REQUISITOS PREVIOS

Conocimientos básicos de Python. Conocimientos básicos de estadística o realización del curso "Estadística, dominando los fundamentos" (Ref. STA).

### COMPETENCIAS DEL FORMADOR

Los expertos que imparten la formación son especialistas en las materias tratadas. Han sido validados por nuestros equipos pedagógicos, tanto en el plano de los conocimientos profesionales como en el de la pedagogía, para cada curso que imparten. Cuentan al menos con entre cinco y diez años de experiencia en su área y ocupan o han ocupado puestos de responsabilidad en empresas.

### MODALIDADES DE EVALUACIÓN

El formador evalúa los progresos pedagógicos del participante a lo largo de toda la formación mediante preguntas de opción múltiple, escenificaciones de situaciones, trabajos prácticos, etc. El participante también completará una prueba de posicionamiento previo y posterior para validar las competencias adquiridas.

### MEDIOS PEDAGÓGICOS Y TÉCNICOS

- Los medios pedagógicos y los métodos de enseñanza utilizados son principalmente: ayudas audiovisuales, documentación y soporte de cursos, ejercicios prácticos de aplicación y ejercicios corregidos para los cursillos prácticos, estudios de casos o presentación de casos reales para los seminarios de formación.
- Al final de cada cursillo o seminario, ORSYS facilita a los participantes un cuestionario de evaluación del curso que analizarán luego nuestros equipos pedagógicos.
- Al final de la formación se entrega una hoja de presencia por cada media jornada de presencia, así como un certificado de fin de formación si el alumno ha asistido a la totalidad de la sesión.

### MODALIDADES Y PLAZOS DE ACCESO

La inscripción debe estar finalizada 24 horas antes del inicio de la formación.

### ACCESIBILIDAD DE LAS PERSONAS CON DISCAPACIDAD

¿Tiene alguna necesidad específica de accesibilidad? Póngase en contacto con la Sra. FOSSE, interlocutora sobre discapacidad, en la siguiente dirección psh-accueil@orsys.fr para estudiar de la mejor forma posible su solicitud y su viabilidad.

#### 4) Algoritmos no supervisados

- Agrupación jerárquica.
- Agrupación no jerárquica.
- Enfoques mixtos.

*Trabajo práctico* : Procesamiento no supervisado de agrupaciones en varios conjuntos de datos.

#### 5) Análisis de componentes

- Análisis de componentes principales.
- Análisis factorial por correspondencia.
- Análisis de correspondencias múltiples.
- Análisis factorial para datos mixtos.
- Clasificación jerárquica basada en componentes principales.

*Trabajo práctico* : Aplicación de la reducción del número de variables e identificación de los factores subyacentes a las dimensiones asociadas a una variabilidad significativa.

#### 6) Análisis de datos de texto

- Recogida y tratamiento previo de datos textuales.
- Extracción de entidades primarias, entidades con nombre y resolución referencial.
- Etiquetado gramatical, análisis sintáctico, análisis semántico.
- Lematización.
- Representación vectorial de textos.
- Ponderación TF-IDF.
- Word2Vec.

*Trabajo práctico* : Explore el contenido de una base de datos de texto mediante el análisis semántico latente.

## FECHAS

---

Contacto