



Introducción a la ciencia de datos utilizando PYTHON

ÁREA: DESARROLLO DE SOFTWARE

PRESENTACIÓN

Diariamente en Internet se genera y almacena información equivalente a 2 Exabytes, con una tendencia al alza cada año. Cada minuto se suben 72 horas de video en YouTube®, más de 500,000 mensajes en Twitter®, alrededor de 100,000 nuevos artículos en sitios como WordPress® y más de 10,000 fotografías en Instagram® y Flickr®. La Ciencia de Datos se emplea en este entorno para establecer predicciones y tomar decisiones en prácticamente cualquier ámbito donde haya grandes volúmenes de datos.

PERFIL DE INGRESO

Este curso está dirigido a las personas interesadas en el uso de las herramientas de la Ciencia de Datos, para establecer análisis predictivos y retrospectivos. Se requiere haber acreditado o demostrar conocimientos equivalentes al curso Introducción a la programación.

OBJETIVO

El participante reconocerá las características básicas del flujo de trabajo para el análisis de datos, empleando las herramientas estadísticas y computacionales para su uso elemental en la Ciencia de Datos.

TEMARIO

1. INTRODUCCIÓN
 - 1.1 Conceptos de Machine Learning
 - 1.2 La ciencia de datos
 - 1.3 Flujo de trabajo del análisis de datos
 - 1.4 Aplicaciones de la Ciencia de Datos
2. CONFIGURACIÓN DEL ENTORNO DE TRABAJO PARA EL ANÁLISIS DE DATOS
 - 2.1 Instalación de las herramientas de programación Python
 - 2.2 Instalación de paquetes de análisis de datos (pip, PyCharm, Anaconda y Jupyter)
 - 2.3 Instalación del IDE de programación
3. INTRODUCCIÓN AL LENGUAJE DE PROGRAMACIÓN PYTHON
 - 3.1 Estructura y elementos del lenguaje
 - 3.2 Módulos, paquetes y espacios de nombre

- 3.3 Programación orientada a objetos
- 3.4 Uso de bibliotecas
- 4. INTRODUCCIÓN A LOS ELEMENTOS DE ANÁLISIS ESTADÍSTICO PARA EL ANÁLISIS PREDICTIVO Y EL ANÁLISIS RETROSPECTIVO
 - 4.1 Conceptos básicos de Estadística
 - 4.2 Regresión Lineal con Python
 - 4.3 Regresión Logística con Python
 - 4.4 Segmentación o Clustering
 - 4.5 Árboles de decisión
 - 4.6 Series temporales
 - 4.7 Técnicas de clasificación y predicción supervisada Naïve Bayes
 - 4.8 Procesamiento de Lenguaje Natural
- 5. OPERACIONES DE PROCESAMIENTOS DE DATOS
 - 5.1 Proceso de preparación de datos o Data Cleaning
 - 5.2 Proceso de limpieza, estructuración y enriquecimiento de datos o Data Wrangling
 - 5.3 Clustering y Clasificación
 - 5.3.1 Árboles y bosques aleatorios
 - 5.3.2 Algoritmo estadístico de clasificación supervisado: K nearest neighbor
 - 5.3.3 Elementos básicos de redes neuronales
- 6. Perspectivas de la Ciencia de Datos
 - 6.1 El futuro de la Ciencia de Datos

PERFIL DE EGRESO

Los conocimientos adquiridos permitirán al participante conocer algunas de las herramientas empleadas para el tratamiento de análisis estadístico, así como las operaciones de procesamiento de datos involucradas en la Ciencia de Datos.

REQUISITOS ACADÉMICOS

Estudios mínimos de preparatoria o equivalente.

DURACIÓN

30 horas

RECURSOS INFORMÁTICOS

- Python Pip
- PyCharm
- Anaconda y Jupyter
- conexión a Internet

BIBLIOGRAFÍA

- Deitel, Paul J. and Deitel, Harvey M. Intro to Python for Computer Science and Data Science: Learning to Program with AI, Big Data and the Cloud. Pearson, 2019, 880 pp.
- Provost, Foster and Fawcett, Tom. Data Science for Business. What You Need to Know about Data Mining and Data-Analytic Thinking. O'Reilly, 2013, 414 pp.

- Downey, Allen. Elements of Data Science. 2022. En: <https://alldowney.github.io/ElementsOfDataScience/README.html>. Fecha de consulta: 20 de octubre de 2022.
- Downey, Allen. Think Stats 2e. Green Tea Press - Free books by Allen B. Downey. 2022. En: <https://greenteapress.com/wp/think-stats-2e/>. Fecha de consulta: 20 de octubre de 2022.
- Downey, Allen. Think Python 2e. Green Tea Press - Free books by Allen B. Downey. 2022. En: <https://greenteapress.com/wp/think-python-2e/>. Fecha de consulta: 20 de octubre de 2022.
- Downey, Allen. Think Bayes 2e. Green Tea Press-Free books by Allen B. Downey. 2021. En: <https://greenteapress.com/wp/think-bayes/>. Fecha de consulta: 20 de octubre de 2022.

