



Programa Iberoamericano de
Formación en Minería de Datos



Programa Iberoamericano de
Formación en Minería de Datos

Ciencia de Datos con Python



(506) 4030.1205 - (506) 4030.1114



info@promidat.com



facebook.com/promidat



Twitter.com/promidat



www.promidat.com

Tutor: El curso será impartido por el Dr. Oldemar Rodriguez Rojas. El Dr Rodríguez realizó sus estudios doctorales en la Universidad de París IX Dauphine, Francia y un Posdoctorado en Ciencias de Datos de la Universidad de Stanford California. Es Catedrático de la Universidad de Costa Rica y Director de Formación de PROMiDAT. Premio Nacional de Ciencia y Tecnología Clodomiro Picado Twight en el año 2001. Por sus trabajos y patentes en modelos predictivos simbólicos para predecir fraudes en tarjetas de crédito recibió la condecoración de la Medalla de Oro para el Inventor Destacado del año 2004 por la Organización Mundial de Propiedad Intelectual (OMPI) con sede en Ginebra Suiza.

Duración: 8 semanas.

Descripción:



En este curso se estudiará en primera instancia a Python como un Lenguaje de Programación Funcional y Orientado a Objetos.

Luego se estudiarán en detalle la implementación de Métodos Exploratorios y de Clustering (Aprendizaje no Supervisado) en Python, métodos como Análisis en Componentes Principales, Agrupación Jerárquica y k-medias serán estudiados.

También se estudiarán en detalle la implementación de métodos Métodos Predictivos y de Regresión (Aprendizaje Supervisado) en Python, métodos como los K vecinos más cercanos, el Método de Bayes, Análisis Discriminante Lineal y Cuadrático, Máquinas Vectoriales de Soporte, Árboles de Decisión, Bosques Aleatorios (Random Forest) y Métodos de Potenciación (Boosting)



Objetivos: El estudiante será capaz de:

1. Programar en Lenguaje Python con una filosofía Funcional o desde una perspectiva Orientada a Objetos.
2. Utilizar Python para generar modelos exploratorios (aprendizaje no supervisado).
3. Utilizar Python para generar modelos predictivos (aprendizaje supervisado).
4. Analizar tablas de datos reales utilizando scrips propios y paquetes especializados de Python.

Metodología:

Basado en la teoría y en la aplicación directa de los conceptos aprendidos. Para esto se dispondrán de las siguientes herramientas.

- Una vídeo conferencia semanal, las cuales quedarán grabadas en Webex, para que los alumnos la puedan acceder en cualquier momento desde el Aula Virtual.
- Trabajos prácticos semanales.
- Foros para plantear dudas al tutor y compañeros.
- Aula virtual en Moodle.

Contenido:

1. Python como Lenguaje de Programación
 - a) Python 2 versus Python 3
 - b) Instalando Anaconda, Spyder, Scikit-Learn, NumPy, SciPy, IPython, Jupiter, Matplotlib, Pandas y Sympy.
 - c) Variables y tipos de datos
 - d) Listas y Diccionarios en Python
 - e) Sentencias de control tipo "if"
 - f) Ciclos tipo "While"
 - g) Funciones en Python
 - h) Definiciones en Orientación Orientada a Objetos
 - i) Diseño de programas Orientados a Objetos

- j) Objetos en Python
- k) Herencia simple y múltiple
- l) Clases Abstractas y Polimorfismo
- 2. Python en Análisis de Datos Exploratorio
 - a) Uso de NumPy
 - b) Manipulación de Datos con Pandas
 - c) Visualización de Datos con Matplotlib
- 3. Aprendizaje no supervisado en Python (Clustering)
 - a) Introducción a Scikit-Learn
 - a) Análisis en Componentes Principales
 - b) Clustering Jerárquico
 - c) El método de K-medias
- 4. Aprendizaje supervisado en Python (Métodos Predictivos)
 - a) El método de los K vecinos más cercanos
 - b) Método de Bayes
 - c) Análisis Discriminante Lineal y Cuadrático
 - d) Máquinas Vectoriales de Soporte
 - e) Árboles de Decisión
 - f) Bosques Aleatorios (Random Forest)
 - g) Métodos de Potenciación (Boosting)
 - h) Redes Neuronales
 - i) Validación cruzada en Scikit-Learn

Bibliografía:

1. Andreas C. Müller and Sarah Guido. Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly, 1st Edition, 2017.



2. Dusty Phillips. Python 3 Object-oriented Programming, Second Edition. Packt Publishing Ltd, 2015.
3. Eric Matthes. Python Crash Course A Hands-On, Project-Based introduction to Programming. No Starch Press, Inc. 2016.
4. Jake VanderPlas. Python Data Science. O'Reilly, 2017.
5. John Paul Mueller (Author) and Luca Massaron. Python for Data Science For Dummies (For Dummies (Computer/Tech)) 1st Edition, 2015.
6. Steven F. Lott. Mastering Object-oriented Python. Packt Publishing Ltd, 2014.
7. Python Software Foundation. 2017. [Python 3.6.2 documentation](https://docs.python.org/3.6.2/). python.org.
8. Anaconda 2017. [Download Anaconda Distribution Python 3.6 version](https://www.anaconda.com/distribution/#python36). Anaconda Inc.
9. Anaconda 2017. [Anaconda Documentation](https://docs.anaconda.com/). Anaconda Inc.