

MÓDULO: 1 – LAS HERRAMIENTAS DEL CIENTÍFICO DE DATOS PROFESOR: JUAN MANUEL MORENO LAMPARERO

Descripción del Módulo

En el mundo en el que vivimos cada vez es más frecuente oír hablar de términos como el Big Data, la Inteligencia Artificial (IA), el internet de las cosas (IoT), la industria 4.0, etc. Todos estos ámbitos tienen un gran punto en común la **ciencia de datos** y es que, la sociedad y las empresas están inmersas en un mundo cada vez más digital y, más conectado en el que se producen ingentes cantidades de datos que deben ser tratados para extraer valor sobre los mismos y mejorar la toma de decisiones.

Para este procesamiento de datos existen una infinidad de herramientas que integran diferentes módulos para tomar y persistir datos en un punto de origen, realizar un análisis exploratorio de los mismos y, aplicar algoritmos con el fin de automatizar procesos u obtener predicciones que sirvan de base para la optimización de procesos empresariales y por supuesto, mejorar los beneficios. Dada la gran variedad de herramientas disponibles en un mercado cada vez más competitivo, en este módulo el alumno comprenderá la base que hay detrás de esas herramientas a través del desarrollo software en los dos lenguajes más populares de ciencia de datos en la actualidad que son Python y R.

Objetivos del Módulo

- Conocer las principales herramientas para la ciencia de datos y cómo ponerlas en práctica en cualquier sector.
- Entender y comprender los fundamentos de los lenguajes de programación: Python y R.
- Comprender cómo utilizar las estructuras de datos más utilizadas en ciencia de datos o Data Science.
- Aprender a instalar y utilizar paquetes o librerías para extender las funcionalidades de un lenguaje de programación.
- Saber comunicar a través de visualización, resultados y transformaciones realizadas sobre las fuentes de datos.
- Aprender cuáles son algunos de los pasos fundamentales de un análisis exploratorio de datos.

Metodología

Este módulo será eminentemente práctico, ya que se utilizarán herramientas de desarrollo software prácticamente en todo momento, no obstante, el alumno contará por una parte de todos los apuntes en la plataforma de la escuela, presentaciones en las que se muestre la parte fundamental y objetivos de cada unidad y, finalmente archivos en los que podrá encontrar los códigos necesarios para aprender todos los aspectos de programación que se impartirán sobre Python y R.

En todas las sesiones habrá diferentes ejercicios disponibles para reforzar las competencias que se estén trabajando tanto en la sesión actual, como en las sesiones anteriores.

Una vez visto todo el contenido del módulo, para encaminar al alumno hacia la práctica final se ofrecerán dos casos prácticos resueltos, uno en Python y otro en R, trabajando en todo momento con competencias vistas en las unidades.

La evaluación del módulo se realizará mediante un caso práctico que constará de dos ejercicios, en los que se le solicitará al alumno un análisis exploratorio de datos, uno en Python y otro, en R.

Contenidos (36 horas)

Tema 1. Fundamentos de Python:

- El lenguaje de programación Python.
- Entorno de programación – Jupyter Notebook.
- Sintaxis básica de Python.
- Herramientas para control de flujo.
- Estructuras de datos.
- Funciones.

Tema 2. Procesamiento de datos y librerías para ciencia de datos Numpy y Pandas:

- Procesamiento de archivos.
- Numpy.
- Pandas.

Tema 3. Procesamiento de datos y visualización con Python:

- Matplotlib.
- Seaborn
- Plotly
- Plotnine - Ggplot (Opcional).
- Bokeh (Opcional).

Tema 4. Fundamentos de R:

- El lenguaje de programación R.
- Entorno de programación – RStudio.
- Sintaxis básica de R.
- Estructuras de datos.
- Control de flujo.
- Funciones.

Tema 5. Paquetes de R:

- Instalación y carga de paquetes.
- Procesamiento de datos con Dplyr.
- Procesamiento de datos con TidyR.

Tema 6. Procesamiento de datos y visualización con R:

- Gráficos básicos en R.
- Gráficos en capas Ggplot2.
- Gráficos dinámicos Plotly.

MÁSTER EN BUSINESS ANALYTICS & DATA SCIENCE

CURSO 2023-2024



Evaluación de la asignatura

ACTIVIDAD	PONDERACIÓN
Caso práctico final: - Anlisis exploratorio de datos (EDA) en Python. - Anlisis exploratorio de datos (EDA) en R.	100 % (50% Python, 50% R)

Requerimientos Técnicos

Indícanos los requerimientos técnicos que vas a necesitar a lo largo del módulo, así como algún tipo de software específico.

Anaconda Distribution https://www.anaconda.com/products/distribution
R base https://cran.r-project.org/
R Studio https://posit.co/products/open-source/rstudio/

Bibliografía

- Dr. Charles Russell Severance. Python for Everybody: Exploring Data in Python 3. CreateSpace Independent Publishing Platform; 2011.
- Python para todos. Explorando la información con Python 3. Charles R. Severance https://do1.dr-chuck.com/pythonlearn/ES_es/pythonlearn.pdf
- Allen B. Downey. Think Python: How to Think Like a Computer Scientist. O'Reilly Media; 2011
- Wes McKinney. "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython". 2017; O'Reilly UK Ltd
- Michael Heydt. "Learning pandas – Python Data Discovery and Analysis Made Easy", Packt Publishing; 2015
- Ivan Idris. "Python Data Analysis Cookbook". Packt Publishing; 2016

MÁSTER EN BUSINESS ANALYTICS & DATA SCIENCE

CURSO 2023-2024



- Aldrin Yim, "Matplotlib for Python Developers Effective techniques for data visualization with Python, 2nd Edition". Packt Publishing ; 2018
- Kevin Jolly, "Hands-On Data Visualization with Bokeh Interactive web plotting for Python using Bokeh". Packt Publishing ; 2018
- Gergely Daróczi. "Mastering Data Analysis with R". 2015; Pack Publishing.
- Julio Sergio Santana "El arte de programar en R. Un lenguaje para la estadística". 2014
- Michael J. Crawley. "The R Book". 2013; John Wiley & Sons, Ltd
- Garrett Golemund. "R for Data Science". 2017; O'Reilly
- Bill Yarberry. "DPLYR: In One Hour Learn Powerful, Practical Data Munging Techniques. Take Your R Skill to the Next Level". 2018; Independently published
- Bradley C. Boehmke. "Data Wrangling with R". 2016; Springer
- Winston Chang. "R Graphics Cookbook: Practical Recipes for Visualizing Data". 2013; O'Reilly Media