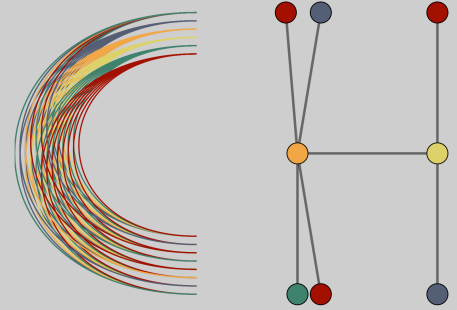


HERRAMIENTAS DE ANÁLISIS DE DATOS

Curso breve de técnicas modernas de análisis

De una forma o de otra, es común enfrentarse con datos. Este curso pretende ser una introducción de las diversas herramientas disponibles para su análisis. Se cubrirá programación básica-intermedia en R y en Python, así como buenas prácticas para controlar versiones de código, datos, y modelos.



TEMARIO

- 2 horas
 - Introducción**
 - Flujos de trabajo y pipelines
 - Lenguajes de programación
 - Editores de texto e IDEs
 - Control de versiones
 - Ambientes productivos
 - Dataframes vs. databases
- 12 horas
 - R**
 - Editores e IDEs | *RStudio*, *VS Code*, *Jupyter*
 - APIs de manejo de datos | *dplyr*, *data.table*, *base*
 - Visualización de datos | *ggplot2*, *ggforce*, *ggraph*
 - Modelos | *tidymodels*
 - Comunicación de resultados | *RMarkdown*, *Shiny*
- 12 horas
 - Python**
 - Editores e IDEs | *Jupyter*, *VS Code*
 - APIs de manejo de datos | *pandas*
 - Visualización de datos | *seaborn*, *matplotlib*
 - Modelos | *scikit-learn*
 - Comunicación de resultados | *Jupyter*
- 5 horas
 - Control de versiones**
 - Código | *git*, *GitHub*
 - Datos | *DVC*
 - Modelos | *MLflow*
- A disp. de tiempo
 - Temas adicionales**
 - Makefiles | *GNU Makefiles*
 - Modelos en producción | *Docker*, *Flask*, *Kubernetes*, *Spark*, *unittest*, *testthis*, *plumber*
 - Análisis de redes | *Neo4j*



INFORMACIÓN ADICIONAL

- Audiencia**
 - Cualquier persona con interés o necesidad de trabajar con datos.
 - Cada sesión comenzará desde cero e irá construyendo hacia ejemplos de mayor complejidad. Sin embargo, dada la disponibilidad de tiempo, el ritmo será acelerado.

CONTACTO

✉ haro_ca@outlook.com
🐦 [@haro_ca](https://twitter.com/haro_ca)
🐙 github.com/haro-ca
📖 medium.com/@haro_ca

OBJETIVO

El objetivo del curso es proporcionar un panorama completo del ecosistema de trabajo con datos. Se expondrán los principales retos que es común enfrentar, así como una introducción a las herramientas necesarias para poder entregar un producto completo. Los temas cubiertos son, a mi consideración, los más importantes (e interesantes) para llevar a cabo un análisis de datos reproducible y en línea con buenas prácticas del ámbito.

El curso no será suficiente para dar un conocimiento profundo, su intención es desarrollar un lenguaje mínimo para comenzar a practicar, indagar, y profundizar en el trabajo con datos.

En ese sentido, este temario, las sesiones, y el material adicional, fungen como un compendio y una galería de técnicas.

Este temario fue realizado en R con *page-down*.

Última actualización 2020-02-29.

Requisitos de software

Se usará R, Python, SQL y Bash para demostrar el uso de RStudio, Jupyter Lab/Notebook, MLflow, git, MLflow, DVB, PostgreSQL, GNU Makefiles, Docker, Flask, Kubernetes, Spark.

Para el material adicional, se necesitará la instalación de todas las herramientas.

Para las sesiones se harán ejemplos rápidos de la mayor parte de ellas, así que es posible solo instalar R, Python, git, RStudio, Jupyter Lab/Notebook y VS Code.

Conocimientos previos

No se necesita ningún conocimiento previo de las herramientas de software. pero sí será necesario conocimiento inferencia estadística para las secciones de modelos.

Material

Toda será publicado en github.com/haro-ca/teaching/heramientas_iniciales, tanto código cubierto, como ejercicios y ejemplos adicionales.



REFERENCIAS

Libros

H. Wickham, G. Grolemund. "R for Data Science". <https://r4ds.had.co.nz/>

Hadley Wickam. "Advanced R". <https://adv-r.hadley.nz/>

Yihui Xie, J.J. Allaire, Garrett Grolemund. "R Markdown".

<https://bookdown.org/>

Wes McKinney. "Python for data analysis".

Richard McElreath. "Statistical Rethinking".

Yifan Wu. "Is a dataframe just a table?". <http://yifanwu.net/>

Blogs

Vincent Warmerdam. <https://koaning.io> y <https://calmcode.io>

Tom Augspurger. <https://tomaugspurger.github.io/>

Alex Ionnides. <https://alexioannides.com/>

Julia Silge. <https://juliasilge.com/>

Thomas P. Lindersen. <https://www.data-imaginist.com/>

Arturo Gonzales Bencomo.

<https://www.medium.com/@arturo102964>