

Análisis Masivo de Datos I

MSc. Alvaro Limber Chirino

Unidades de aprendizaje

1. Introducción al ecosistema de R
2. Introducción al Big Data
3. Fuentes de datos
4. Gramática de datos
5. Muestreo y filtrado
6. Modelado con software con MapReduce

Evaluación

Formal

- 20% Laboratorio 1: Temas 1 al 3
- 20% Laboratorio 2: Tema 4
- 20% Laboratorio 3: Temas 5 y 6
- 20% Examen teórico: Todos los temas
- 20% Proyecto Dashboard (Shiny/grupal)

Extra:

- 10% **Poster estadístico:**
 - Visto Bueno hasta el 2 de marzo
 - Más información: <https://islp-bolivia.github.io/>
- 10% **Podcast (grupal):** Estado, limitaciones y desafíos del Big Data en Bolivia
 - Entrega el día del examen teórico
 - Entregables: audio, guion e informe corto en rmd.
- 5% **Dialogo Socrático:** Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World
 - Leer los primeros 5 capítulos
 - Actividad conjunta a la presentación del dashboard

Planificación por fecha

Sesiones	TEMA	Actividad	Fecha
1	0	Preparación de la materia	2023-02-02
2	1	Herramientas y canales de comunicación (GIT)	2023-02-07
3	1	Introducción a R e investigación reproducible (rmd)	2023-02-09
4	1	Estructuras homogéneas y heterogéneas	2023-02-14
5	1	Loops, condiciones y funciones	2023-02-16
6	1	Dashboard en R	2023-02-21
7	1	Shiny	2023-02-23
8	2	Introducción al Big Data	2023-02-28
9	2	Conceptos y definiciones	2023-03-02
10	2	Flujo de trabajo e inferencia	2023-03-07
11	2	Calidad del dato, captura y preservación	2023-03-09
12	2	Análisis, modelado, inferencia y ética	2023-03-14
13	2	Almacenamiento en Big data	2023-03-16
14	3	Fuentes de datos y enfoques	2023-03-21
15	3	Centro de datos importación de datos, ETL, ELT	2023-03-23
16	3	Raspado web 1	2023-03-28
17	3	Raspado web 2	2023-03-30
18	3	API 1	2023-04-04
19	3	API 2	2023-04-06
20	0	Laboratorio 1	2023-04-11
21	4	Introducción a la gramática de datos	2023-04-13
22	4	Tidyverse	2023-04-18
23	4	Filtrado y selección	2023-04-20
24	4	Transformación y reportes	2023-04-25
25	4	Estadística descriptiva	2023-04-27
26	4	Inferencia descriptiva	2023-05-02
27	4	Visualización de datos	2023-05-04
28	0	Laboratorio 2	2023-05-09
29	5	Usos del muestreo en big data	2023-05-11
30	5	Tipos de muestreo	2023-05-16
31	5	Selección de la muestra y estimación	2023-05-18
32	6	Introducción e instalación de spark	2023-05-23
33	6	Importación de datos	2023-05-25
34	6	Análisis básico	2023-05-30
35	0	Laboratorio 3	2023-06-01
36	0	Examen teórico	2023-06-06
37	0	Presentación de dashboard / Dialogo socrático	2023-06-08
38	0	Sistemación de notas	2023-06-13
39	0	Entrega y revisión de notas	2023-06-15
40	0	Notas al sistema	2023-06-20

Referencias

- Erl, Thomas, Wajid Khattak, and Paul Buhler. 2016. *Big Data Fundamentals Concepts, Drivers & Techniques*.
- Iacus, Stefano M. 2015. *Automated Data Collection with r - a Practical Guide to Web Scraping and Text Mining*. *Journal of Statistical Software*. Vol. 68. <https://doi.org/10.18637/jss.v068.b03>.
- . 2017. *Big Data and Social Science - a Practical Guide to Methods and Tools*. *Journal of Statistical Software*. Vol. 78. <https://doi.org/10.18637/jss.v078.b02>.
- Luraschi, J., K. Kuo, E. Ruiz, and an O'Reilly Media Company Safari. 2019. *Mastering Spark with r*. O'Reilly Media, Incorporated. <https://therinspark.com/>.

Wickham, H., and G. Grolemund. 2017. *R for Data Science*. O'Reilly Media. <https://r4ds.had.co.nz/>.