



PROMiDAT

IBEROAMERICANO

Programa Iberoamericano de
Formación en Minería de Datos

CD303

Programación Avanzada en R



(506) 4030.1205 - (506) 4030.1114



info@promidat.com



facebook.com/promidat



Twitter.com/promidat



www.promidat.com

Índice

Duración	2
Descripción	2
Objetivos	2
Metodología	2
Contenido	3
Evaluación	3
Bibliografía	4

Duración

Cuatro semanas.

Descripción



En este curso se presentan desde los elementos básicos hasta la elaboración de un proyecto del proceso de manipulación, visualización de datos y la elaboración de reportes para resolver un problema de Ciencia de Datos en R.

Objetivos

En este curso el estudiante será capaz de:

1. Verificar el formato correcto de los datos y realizar correcciones.
2. Crear resúmenes de datos e interpretarlos.
3. Comprender los conceptos básicos de la visualización de datos.
4. Aplicar las mejores prácticas para la elaboración de gráficos y reportes.
5. Diseñar procesos de análisis de datos replicables.

Metodología

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

1. Una vídeo conferencia semanal, las cuales quedarán grabadas en Zoom, para que los alumnos la puedan acceder en cualquier momento.
2. Trabajos prácticos semanales.
3. Foros para plantear dudas al tutor y compañeros.
4. Aula virtual en Moodle.

Luego de este curso el estudiante será capaz de:

Implementar proyectos de Minería de Datos que involucre los pasos básicos de manipulación, visualización y reportes de nivel empresarial utilizando el lenguaje de programación R.

Contenido

1. **Creación de proyectos.**
 - a. Control de rutas.
 - b. Estructura de carpetas.
 - c. Buenas prácticas para el diseño de proyectos.
2. **Diseño de reportes con Quarto y KnitR.**
 - a. Sintaxis de Quarto y Knit.
 - b. Buenas prácticas para el diseño de reportes.
 - c. Modificar tipos de datos de un archivo.
3. **Manipulación de datos con Tidyverse.**
 - a. Selección de variables.
 - b. Selección de individuos.
 - c. Resúmenes de datos.
 - d. Ordenar tablas.
 - e. Creación de nuevas variables.
4. **Programación de Métodos Predictivos.**
 - a. KNN.
 - b. Árboles.
 - c. Bosques Aleatorios.
 - d. Potenciación y XGBoosting.
5. **Calibración y Selección de Modelos Predictivos desde la consola de R.**
 - a. Calibración y Selección de Modelos de Clasificación.
 - b. Calibración y Selección de Modelos de Regresión.

Evaluación

El curso se evalúa con 4 tareas, una por semana, cada tarea tiene un valor de 25 puntos. La nota mínima de aprobación es de 70.

Bibliografía

- [1] J.J. Allaire et al. *Quarto*. DOI: 10.5281/ZENODO.5960048. Zenodo, Jan. 10, 2022. DOI: 10.5281/ZENODO.5960048. URL: <https://zenodo.org/record/5960048>.
- [2] John M. Chambers. *Software for data analysis: programming with R*. Statistics and computing. OCLC: ocn191243189. New York: Springer, 2008.
- [3] chocolatey. *Chocolatey Software Docs | Chocolatey - Software Management for Windows*. 2022. URL: <https://docs.chocolatey.org/en-us>.
- [4] Colin Gillespie and Robin Lovelace. *Efficient R programming: a practical guide to smarter programming*. First edition. OCLC: ocn964820738. Sebastopol, CA: O’Reilly Media, Inc, 2016.
- [5] Norman S. Matloff. *The art of R programming: tour of statistical software design*. OCLC: ocn711045702. San Francisco: No Starch Press, 2011.
- [6] Hadley Wickham. “Tidy Data”. In: *Journal of Statistical Software* 59 (Sept. 12, 2014), pp. 1–23. DOI: 10.18637/jss.v059.i10. URL: <https://doi.org/10.18637/jss.v059.i10>.
- [7] Hadley Wickham and Garrett Grolemund. *R for data science: import, tidy, transform, visualize, and model data*. First edition. OCLC: ocn968213225. Sebastopol, CA: O’Reilly, 2016.