



PROMiDAT

IBEROAMERICANO

Programa Iberoamericano de
Formación en Minería de Datos

Programación en Lenguaje R



(506) 2268.8823 - (506) 8708.9091



info@promidat.com



facebook.com/oldemarrodriguez



www.promidat.com

Tutor: El curso será impartido por el Dr Oldemar Rodríguez y el Ing. Carlos Agüero.

Duración: Cuatro semanas.

Descripción:



Este curso ofrece una introducción práctica a las técnicas modernas de computación de datos utilizando el lenguaje de programación R. Para esto se estudiará el uso de R en aplicaciones estadísticas y de minería de datos, así como la creación de gráficos estadísticos y la creación documentos auto reproducibles.

Se enseñarán los fundamentos básico de R como lenguaje de programación de computadoras, especialmente orientado a la programación de aplicaciones Estadísticas y de Minería de Datos.

Se explorarán las conexiones a diversos orígenes de datos incluyendo los formatos más utilizados como son csv, archivos de texto plano, xls,.xlsx.

Se estudiará además la instalación y uso de paquetes en R, estos permiten ampliar las capacidades del lenguaje existiendo actualmente más de 10,000 paquetes disponibles para su instalación, entre los utilizados durante el curso están algunos de los principales paquetes utilizados en la industria como son: ggplot2, FactoMineR, scales, graphics, corplot, entre otros.

Los ejemplos de este curso estarán motivados por problemas reales en el campo. Por lo tanto, los estudiantes adquieren conocimientos de muchas herramientas diferentes que pueden combinarse para resolver problemas reales.

Objetivos:

El objetivo fundamental de este curso es que el estudiante sea capaz de manipular y generar datos con facilidad, para esto será capaz de:

1. Comprender los conceptos básicos de la lógica de programación en lenguaje R.
2. Entender la sintaxis y semántica del lenguaje así como poder identificar errores.
3. Instalar y actualizar de paquetes desde el CRAN.
4. Importar datos desde diversos orígenes.
5. Utilizar las funciones provistas por el lenguaje para efectuar análisis de datos.
6. Desarrollar funciones propias en lenguaje R.

7. Presentar los resultados de los análisis a través de gráficos estadísticos
8. Crear documentos autoreproducibles en diversos formatos como htm, doc y pdf.

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.
- Trabajos prácticos semanales.
- Foros para plantear dudas al tutor y compañeros.
- Aula virtual en Moodle.

Luego de este curso el estudiante será capaz de:

Programar en lenguaje R proyectos de Minería de Datos.

Contenido:

- 1) Introducción al uso R
 - a. La consola y el editor en R y Rstudio
 - b. Uso e instalación de paquetes
 - c. Tipos y estructuras de datos
 - i. Vectores
 - ii. Matrices
 - iii. Listas
 - iv. Hojas de datos
 - d. Subsetting de estructuras de datos
 - e. Funciones más comunes
 - f. Operaciones aritméticas y lógicas
 - g. Lectura de datos
 - h. Exportar datos
- 2) Visualización de datos
 - a. Paquetes disponibles para la visualización de datos
 - b. Conceptos básicos de la visualización de datos
 - c. Gráficos para el análisis estadístico de datos
 - i. Dispersión
 - ii. Barras
 - iii. Pastel
 - iv. Cajas

- v. Correlación
 - vi. Densidad
 - vii. Histogramas
- 3) Estructuras de control
- i. Operaciones lógicas
 - ii. Estructura IF ELSE
 - iii. Estructura IFELSE
 - iv. Estructura FOR y WHILE
- 4) Creación de funciones
- a. Estructura fundamental de una función en R
 - b. Parametros
 - i. Valores por defecto
 - ii. Captura de valores ausentes
 - c. Cuerpo de la función
 - d. Ciclo de vida de las variables
 - e. Entorno de una variable
 - f. Retorno de uno y multiples valores
 - g. Funciones como parámetros de otra función
 - h. Funciones como retorno de otra función
 - i. Principios de la programación vectorizada
- 5) Creación de documentos autoreproducibles
- a. Sintaxis del lenguaje de etiquetas Rmarkdown
 - b. Formateo de texto
 - c. Uso de imágenes, enlaces y otros recursos externos
 - d. Manejo de código R embebido en el documento
 - e. Manejo de errores, alertas y mensajes
 - f. Parametrización de documentos
 - g. Formatos de salida pdf, doc, html

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. John M. Chambers. Programming with R: Software for Data Analysis. Springer, Stanford University, Palo Alto, 2008.
2. Graham Williams, Data Mining with Rattle and R. Springer, New York, 2011.

3. Owen Jones, Robert Maillardet and Andrew Robinson. Introduction to Scientific Programming and Simulation using R. Chapman & Hall/CRC Taylor & Francis Group, FL. 2009.
4. R Development Core Team. "R: A Programming Environment for Data Analysis and Graphics". The R Project for Statistical Computing, 2010. <http://www.r-project.org/>
5. R Development Core Team. "Writing R Extensions". The R Project for Statistical Computing, 2010. <http://www.r-project.org/>
6. Wickham, H., & Golemund, G. (2017). *R for data science: Import, tidy, transform, visualize and model data*. Sebastopol: OReilly.
7. Matloff, N. (2013). *The art of R programming: A tour of statistical software design*. San Francisco: No Starch Press.