

Modelos predictivos con R

Sílabo

contacto

+51 (1) 282 9524
+51 9 9038 8434

info@perustat.com
<http://www.perustat.com>
fb://perustat



fundamentación

Durante las últimas décadas se ha producido un desarrollo explosivo en las tecnologías de base de datos y la cantidad de datos que es recolectada. Esto ha creado una oportunidad sin precedentes para la Minería de Datos mediante el proceso de descubrimiento, ya sea supervisado o no, de información interesante y útil a partir de repositorios de datos disponibles. La Minería de Datos está relacionada con el análisis, mayormente estadístico, de grandes conjuntos de datos con la finalidad de proporcionar ideas, patrones, modelos descriptivos y predictivos que permitan extraer y generar conocimiento para las organizaciones. Algunas de las tareas más comunes son la clasificación, agrupamiento, descubrimiento de reglas de asociación y patrones de respuestas, detección de anomalías, etc.

El modelamiento predictivo se define como el proceso de desarrollar herramientas matemáticas o modelos, que generen predicciones correctas. Dentro de los distintos programas estadísticos disponibles, **R** proporciona una poderosa plataforma open source para la aplicación de las técnicas más importantes de la analítica predictiva y es actualmente el programa más utilizado debido, principalmente, al hecho de incluir un mayor número de modelos estadísticos avanzados. Las librerías **Rcmdr** (R Commander) y **Rattle** (R Analytical Tool To Learn Easily) adicionan una interfaz gráfica de usuario específicamente diseñada para facilitar la aplicación de las principales técnicas analíticas a los usuarios que no están acostumbrados al entorno de trabajo de este programa.

Este taller tiene como finalidad que los participantes puedan aplicar los modelos predictivos más usados con el programa **R** a través de las interfaces gráficas R Commander y Rattle.

publico objetivo

Analistas e investigadores de mercado. Profesionales de marketing. Público en general que requiera aplicar técnicas analíticas predictivas.

nivel

Intermedio

conocimientos previos

Conocimientos básicos de inferencia estadística y modelamiento estadístico (recomendable)

logros de aprendizaje

Al finalizar este taller, el participante será capaz de aplicar diversas técnicas analíticas predictivas usando el programa estadístico R con las librerías **Rcmdr** y **Rattle**.

De manera específica el participante estará en capacidad de:

- Acceder a datos almacenados en diversos formatos con las librerías **Rcmdr** y **Rattle**.
- Aplicar modelos de predicción numérica (modelos de regresión) y categórica (modelos de clasificación).
- Comparar y evaluar modelos predictivos.
- Presentar de manera efectiva los resultados obtenidos.

contenidos

- Sesión 1** **Introducción y** manejo de datos con R Commander y Rattle
- Modelamiento predictivo: Conceptos básicos. Predicción numérica vs. clasificación.
 - Balance entre predicción e interpretación.
 - ¿Qué es R?
 - Acceso e instalación de paquetes en R.
 - Código y editores de texto: la interfaz gráfica RStudio.
 - Sintaxis básica en R.
 - Documentación y opciones de ayuda.
 - Primeros pasos con R Commander y Rattle. Lectura de Datos.
- Sesión 2** **Modelos de** regresión
- Definición del Modelo de Regresión Lineal Múltiple.
 - Estimación de los parámetros del modelo.
 - Modelos lineales generalizados.
 - Modelo para datos discretos: regresión de Poisson.
- Sesión 3** **Modelos de** clasificación
- Modelos de regresión binaria: enlaces logit (regresión logística) y probit
 - Árboles de clasificación.
 - Árboles por inferencia condicional.
 - Random Forest
- Sesión 4** **Aplicaciones**
- Caso de estudio.
 - Discusión final.

metodología

La metodología del taller es completamente práctica. Cada sección del curso está motivada por un conjunto de datos en particular, de tal forma que el participante gane experiencia trabajando con una amplia variedad de fuentes de datos similares a los que usa en la realidad. Los contenidos del taller están estructurados en 4 sesiones con un total de 8 horas académicas.

evaluación y asistencia

Se otorgará un certificado a nombre de PeruStat Analytics S.A.C. que acredite la participación en el taller.

materiales

Material preparado por el equipo de capacitación con los contenidos del curso el cuál será entregado a los participantes en medios físicos (modalidad presencial) o digitales (modalidad virtual).

referencias

- Gareth, J., Witten, D., Hastie, T. y Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer Verlag.
- Kuhn, M. y Johnson, K. (2013). Applied Predictive Modeling. Springer Verlag.
- Larose, D.T. (2006). Data Mining Methods and Models. Wiley Interscience.
- Ledolter, J. (2013). Data Mining and Business Analytics with R. John Wiley & Sons.
- Nisbet, R., Elder IV, J. y Miner, G. (2009). Handbook of Statistical Analysis and Data Mining Applications. Academic Press.

Ohri, A. (2012). R for Business Analytics. Academic Press. Springer Verlag.

Putler, D. y Krider, R. E. (2012). Customer and Business Analytics: Applied Data Mining for Business Decision Making Using R. Chapman and Hall/CRC.

Williams, G. (2011). Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery. Springer Verlag.

Wu J. y Coggeshall, S. (2012). Foundations of Predictive Analytics. Chapman and Hall/CRC.