

“DESARROLLO DE MODELOS DE MACHINE LEARNING EN R”

1. Propósitos del taller

Este taller tiene como propósito interiorizar al estudiante en el estudio teórico y práctico de diversas herramientas del área de análisis y predicción de datos.

Actualmente las industrias cada vez más están solicitando profesionales con estas competencias, cuyas aplicaciones abarcan diferentes áreas dentro de la ingeniería como Operaciones, Marketing, Finanzas, etc, donde se busca crear valor a partir de los datos con los que se cuenta.

Los estudiantes aprenderán diversos modelos de aprendizaje supervisado y no supervisado, entendiendo las ventajas y desventajas de cada uno, como su aplicabilidad acorde al contexto, además de aprender cómo implementar estos modelos computacionalmente e interpretar sus resultados.

2. Participantes

Este taller va dirigido a profesionales del área de ingeniería, o carreras como psicología, sociología, etc. que se desarrollen laboralmente con tratamiento de datos, o personas que deseen aprender a extraer información de los datos para agregar valor a la toma de decisiones y ser impulsores de la filosofía “data-driven”.

3. Objetivos del taller

Objetivo general

Este taller tiene como propósito interiorizar al estudiante en el estudio teórico y práctico del área de análisis y predicción de datos usando modelos de Machine learning en el lenguaje de programación R

Objetivos específicos

- Conocer distintos modelos de Machine Learning de aprendizaje supervisado y no supervisado.
- Identificar que modelos conviene usar para responder una hipótesis planteada dependiendo del caso de uso a evaluar.
- Implementar computacionalmente los modelos aprendidos en R para obtener resultados e interpretarlos.

4. Contenidos del taller

Este taller está compuesto por 6 sesiones de 1 hora y 30 minutos cada una, cuyos contenidos son los siguientes:

4.1 Sesión 1: Introducción al análisis de datos y primeros pasos en R

En esta sesión se hará una introducción sobre en qué consiste el análisis de datos, el flujo de creación de valor, la metodología KDD, y el “Storytelling” para mostrar los resultados.

Esto es vital a la hora de diseñar el análisis correspondiente al cual aplicaremos los modelos de Machine Learning.

Finalmente haremos una introducción al lenguaje de programación R donde se desarrollarán operaciones y funciones para aprender a ocupar el software, como también la lectura, escritura y análisis de tablas de datos para su posterior procesamiento.

4.2 Sesión 2: Regresiones lineales con y sin Regularización

En esta sesión revisaremos en que consiste el modelo de regresión lineal, sus bases, ventajas y desventajas, y 2 tipos de regularizaciones (penalizaciones) usadas para este modelo: Ridge y LASSO.

Finalmente se llevará a la práctica con unos ejemplos a desarrollar en R, donde aprenderán a aplicar estos modelos de Machine Learning e interpretar sus resultados.

4.3 Sesión 3: Regresión Logística y Multinomial

En esta sesión revisaremos en que consiste el modelo de regresión logística, sus bases, ventajas y desventajas, y una variación llamada Regresión Multinomial.

Finalmente se llevará a la práctica con unos ejemplos a desarrollar en R, donde aprenderán a aplicar estos modelos de Machine Learning e interpretar sus resultados.

4.4 Sesión 4: Árboles de Decisión y Bosques Aleatorios (Random Forest)

En esta sesión revisaremos en qué consisten los modelos de clasificación de árboles de decisión y los Bosques Aleatorios (Random Forest), sus bases, ventajas y desventajas, etc.

Finalmente se llevará a la práctica con unos ejemplos a desarrollar en R, donde aprenderán a aplicar estos modelos de Machine Learning e interpretar sus resultados.

4.5 Sesión 5: Support Vector Machine (SVM) y Support Vector Regression (SVR)

En esta sesión revisaremos en qué consiste el modelo de programación lineal SVM utilizado para problemas multiclases, sus bases, ventajas y desventajas, y una variación (SVR) utilizada para hacer regresiones de series de tiempo.

Finalmente se llevará a la práctica con unos ejemplos a desarrollar en R, donde aprenderán a aplicar estos modelos de Machine Learning e interpretar sus resultados.

4.6 Sesión 6: K-means y Hierarchical Clustering

En esta última sesión se abordarán dos modelos de aprendizaje no supervisado: K-means y Hierarchical Clustering. Revisaremos sus bases como también sus ventajas y desventajas respectivas a la hora de implementarlos.

Finalmente se llevará a la práctica con unos ejemplos a desarrollar en R, donde aprenderán a aplicar estos modelos de Machine Learning e interpretar sus resultados.

5. Metodología

Este taller constará de 6 sesiones de 90 minutos de duración, las cuales se dividirán en dos partes:

- Una primera parte expositiva, donde se explicarán los modelos de Machine Learning que se enseñarán, con el fin de que el alumno pueda identificar en qué casos conviene utilizarlo dado sus ventajas y desventajas respectivas, y que también pueda evaluar la calidad de su propio modelo desarrollado para saber elegir el mejor dentro de su caso estudiado.
- Una segunda parte práctica, donde se explicará cómo implementar los modelos de Machine Learning estudiados en el lenguaje de programación R, interpretar sus resultados y generar visualizaciones de ellos.

Además, el alumno contará con el material de respaldo para que pueda seguir practicando lo aprendido.

6. Requerimientos

En este taller se usará un el lenguaje de programación R y su IDE Rstudio, los cuales se pueden descargar gratuitamente del siguiente link: <https://rstudio.com/products/rstudio/download/#download>

7. Relator

Pablo Lemus. Ingeniero Civil Industrial de la Universidad de Chile, con un Magíster en Gestión de Operaciones de la misma escuela, y actualmente es estudiante de Doctorado en Sistemas de Ingeniería de la misma casa de estudios.

Pablo cuenta con experiencia en instituciones públicas y privadas, desarrollando proyectos de diferentes temáticas de optimización y toma de decisiones impulsadas por datos. Actualmente se desarrolla como Especialista Data Scientist y también como PhD Student Researcher del Centro de Investigación en Inteligencia de Negocios (CEINE) de la Universidad de Chile. Adicionalmente, se desempeña como docente part-time de la Facultad de Ingeniería y Ciencias de la Universidad Diego Portales.