

**SEGUNDA VERSIÓN ESCUELA DE VERANO CEPR:**

**CIENCIA DE DATOS UTILIZANDO “R”**

**PROGRAMA**

Lunes 6 de enero Introducción al software

Martes 7 de enero Visualización de datos en “R”

Miércoles 8 de enero Introducción a técnicas de Machine Learning

Jueves 9 de enero Taller de Generación de proyectos de Ciencia de Datos

* Todas las clases se realizarán entre las 8:30 y 12:40 en la sala A-209 de la Universidad Adolfo Ibáñez, Campus Viña del Mar.

**CONTENIDOS**

**Introducción al software**: Entender la estructura de las fuentes de información y el cómo trabajarlas es un requisito esencial para poder introducirse a la ciencia de datos. En este tópico se estudiarán las diferentes estructuras de las bases de datos y el cómo alterarlas de forma eficiente para realizar ciencia de datos.

**Visualización de bases de datos:** El uso de visualizaciones es una de las formas más eficaces para comunicar la información de las bases de datos. En este tópico se introducirá el uso de paquetes de visualización de datos en “R” como ggplot2, ggvis, plotly, entre otros.

**Introducción a técnicas de Machine Learning:** El uso de algoritmos para realizar para prevenir comportamientos no deseados es una práctica cada vez más solicitada por organismos privados y públicos. En este tópico se introducirá el uso de modelos de elección discreta, análisis de clúster y árboles de decisión de forma aplicada utilizando paquetes estadísticos del software “R”.

**Taller de Generación de proyectos de Ciencia de Datos:** Este taller presentará una metodología para explorar qué problemas públicos se pueden resolver con ciencia de datos. Se realizará un ejercicio práctico de análisis de factibilidad de un proyecto y se presentará una matriz de madurez de datos.