



UTN.BA

UNIVERSIDAD TECNOLÓGICA NACIONAL
FACULTAD REGIONAL BUENOS AIRES

Clase 1: INTRODUCCIÓN A MACHINE LEARNING Y R

¿Qué es Machine Learning?

Diferencias y similitudes entre Estadística, Machine Learning y Deep Learning (Inteligencia Artificial).

Mercado laboral y ciencia de datos. Perfiles y skills requeridos.

Comparativa de R y Python.

Primeros pasos programando en R con el entorno RStudio.

Filosofía de R para manejo de datos y sus tipos (vectores, data frames)

Práctica en R: Carga de datos y operaciones básicas.

Clase 2: ANÁLISIS EXPLORATORIO DE DATOS EN R

¿Qué es el análisis exploratorio de datos? ¿Qué sucede si no se hace?

Análisis de variables categóricas y numéricas, métricas y gráficos.

Introducción a métricas de dispersión y percentiles.

Visualización y entendimiento de los datos. Práctica en R.

Clase 3: PREPARACIÓN DE DATOS EN R (PARTE 1)

Introducción a los paquetes standard de gráficos y procesamiento de datos:

“tidyverse” con “dplyr”.

Breve introducción a ggplot2.

Cómo filtrar, ordenar, agrupar y crear nuevas columnas. Joins entre tablas.

Práctica en R.

Tipos de gráficos de acuerdo a la necesidad a resolver.

Cómo descubrir nueva información creando variables con criterios de

modelado predictivo.

Clase 4: PREPARACIÓN DE DATOS EN R (PARTE 2)

Introducción conceptual y práctica al análisis de valores nulos (missing) y extremos (outliers). Ejemplos en R.

¿Cómo reemplazar valores nulos y outliers para maximizar el poder predictivo? (básico). Ejemplos en R.

Calcular métricas por grupo, aplicando varias funciones (group by).

Análisis de correlación de variables, clásico R², sus ventajas y sus problemas con los outliers.

Práctica en R.

Clase 5: CREACIÓN DEL MODELO PREDICTIVO (INTRODUCCIÓN).

Análisis de correlación de variables, clásico R², sus ventajas y sus problemas con los outliers.

Análisis de variables respecto a un objetivo dado. Variables en modelos predictivos.

Práctica en R.

Clase 6: CREACIÓN DEL MODELO PREDICTIVO (INTRODUCCIÓN).

¿Cómo identificar las variables más relevantes? Métodos analíticos y gráficos.

¿Hace falta conocer cientos de modelos de machine learning? Explicación de los modelos más populares: Random Forest y Extreme Gradient Boosting Machine (xgboost)

Práctica: Creación y validación de un modelo Random Forest en R.

Introducción a la identificación de errores silenciosos antes de la puesta en producción del modelo.

Práctica: Implementación del modelo Random Forest, predicción de casos nuevos y reporte final de resultados.