

INTRODUCCIÓN A LA CIENCIA DE DATOS

Del dato al modelado estadístico, con aprendizaje automático y reducción de dimensionalidad

DETALLES DEL CURSO:

- Carga Horaria: 4 SEMANAS
- Lenguajes de Programación: R y Python
- Duración Total: 36 Horas Lectivas
- Docente: PhD. Esteban Fernández Tuesta

SEMANA 1: FUNDAMENTOS

Prácticas en R y Python: Visualización con ggplot2

Objetivo: Transformación de datos brutos en información estructurada.

- Estructuras de datos: Tipos, limpieza y preparación inicial.
- Visualización: Construcción de tablas y gráficos interactivos.
- Análisis multivariado: Medidas de resumen, dispersión y asociación.

SEMANA 2: EXPLORACIÓN Y MODELADO ESTADÍSTICO

Modelado Estadístico & Predicción

Objetivo: Formulación, interpretación y validación de modelos continuos y discretos.

- Regresión lineal: Modelado simple, múltiple y ajuste.
- Diagnóstico: Validación rigurosa de supuestos teóricos.
- Regresión logística: Clasificación binaria y probabilidades.

SEMANA 3: APRENDIZAJE SUPERVISADO: CLASIFICACIÓN Y MODELOS FLEXIBLES

Métricas de evaluación:

Matriz de confusión, Sensibilidad, Especificidad, Curva ROC y Área Bajo la Curva (AUC)

- Modelos Clásicos
- Árboles y Ensamblés
- Regularización.

SEMANA 4: APRENDIZAJE NO SUPERVISADO Y PROYECTO INTEGRADOR

Objetivo: Portafolio práctico de Data Science

- Estructura Oculta en los Datos
- Reducción de Dimensionalidad
- APLICACIÓN PRÁCTICA: Proyecto Integrador