

## Técnicas de Machine Learning para Análisis de Datos

### Tarea 3

#### Aplicación de Naive Bayes, Regresión Logística y Métricas

#### Objetivo general

El objetivo de esta tarea es aplicar Naive Bayes, Regresión Logística y utilizar Métricas para su selección en un dataset provisto.

#### Breve descripción de los datos

El dataset **ECH\_22** (Encuesta continua de hogares) tiene un conjunto de variables que se encuentran definidas en el archivo **DICCIONARIO ECH 2022\_1erSemestre**.

#### Tarea

Teniendo como objetivo la predicción de la variable “acceso\_streaming” y basándose en el código visto en clase, se solicita:

- **Tarea 3.1:** Incorporar nuevas variables independientes de interés (además de las sugeridas en el práctico) de acuerdo a las hipótesis del equipo de trabajo como variables predictoras.
- **Tarea 3.2:** Completar el análisis exploratorio de datos (AED). Evaluar y llegado el caso implementar transformaciones de variables independientes e incluso interacción.
- **Tarea 3.3:** Discutir los resultados de las métricas entre modelos y al agregar nuevas variables. Analice los diferentes resultados y seleccione el modelo y la métrica que considere obtiene mejores resultados en test. ¿Qué métrica seleccionaría de acuerdo a las hipótesis del equipo de trabajo?

#### Recursos

- **ECH\_22.csv:** dataset.
- **DICCIONARIO ECH 2022\_1erSemestrexlsx:** diccionario de datos.
- **pr\_4.ipynb, pr\_5.ipynb:** Jupyter notebook de ejemplo.

**Entrega**

Se deberá entregar un jupyter notebook como informe final de la tarea. En el mismo se pretende que se detalle el código utilizado para resolver lo solicitado en la tarea, así como conclusiones o información teórica relevante.

Se deberán detallar los integrantes del equipo (Nombre, Apellido, Número de estudiante).