Sofía Alonso, Delfina Andreucetti y Victoria de León

**Informe de Adult Census Income**

* **Exploración de datos**

Con el data set de adult census income, que contiene datos sobre el. Tiene datos como: edad, educación, clase social, situación matrimonial, país de procedencia, entre otros. Nuestro objetivo era encontrar las personas que tengan un sueldo superior o inferior a 50k.

Para entrenar este modelo, ignoramos distintas columnas para agilizar la respuesta, como peso final, educación, relaciones y estado civil. Además, de cambiar a binario el sexo femenino y masculino para facilitar el procesamiento de datos, de esta manera obtuvimos los siguientes resultados.

* **Pre-procesamiento de datos con justificación**

El flujo de trabajo incluye las siguientes etapas de preprocesamiento de datos: discretización de variables continuas mediante intervalos iguales, imputación de valores faltantes utilizamos el promedio, estandarización de características, y eliminación de características con demasiados valores faltantes o ceros. Se aplica análisis de componentes principales y se seleccionan las características más relevantes. Además, se permite la selección aleatoria de un número fijo de características.

* **Comparación entre modelos e interpretación**

Para visualizar los datos, utilizamos 3 modelos: árbol, regresión logística y neural network. El que mejor resultado obtuvo fue neural network, con 0.819 que es el número más cercano en todas las métricas.

