ENTREGA 1

Luis F Sanchez – Andrés C. Zúñiga – Sindy C. Cavadia

A partir de un dataset sintético generado por un CTGAN que originalmente se basó en otro dataset que buscaba predecir las cantidades de unas reclamaciones de seguros, se buscará predecir un "target" binario que tomará datos de 19 columnas tipo categóricas y 11 columnas tipo continuas.

El dataset se ha tomado de la plataforma Kaggle y corresponde al "Tabular Playground Series – Mar 2021" (https://www.kaggle.com/competitions/tabular-playground-series-mar-2021), consta de 3 archivos:

- **train.csv**: Con los datos de entrenamiento y la columna target.
- **test.csv**: Con el conjunto de pruebas, a las cuales se les tendrá que hallar el target para cada fila.
- **sample_submission.csv**: Un ejemplo de cómo se debe enviar el archivo.

Métricas

La evaluación, al igual que en la competición correspondiente al dataset, será usando el área debajo de una curva ROC (https://es.wikipedia.org/wiki/Curva_ROC) entre los valores de la probabilidad del target predicho y la del target observado, la meta es que el resultado se acerque a 1.0 de la tasa de verdaderos positivos en la medida de lo posible.