



**Tecnológico
de Monterrey**

**Instituto Tecnológico y de Estudios
Superiores de Monterrey**
Campus Puebla

Inteligencia Artificial para la ciencia de datos TC3007C

**Momento de Retroalimentación: Módulo 1 Utilización,
procesamiento y visualización de grandes volúmenes de datos
(Portafolio Análisis)**

Fernando Jiménez Pereyra A01734609

25 de noviembre de 2022

Decidí utilizar el dataset de [Tabular Playground Series - Oct 2021](#), el cual consiste en datos generados de forma sintética por redes neuronales generativas adversarias. Utilizando un modelo de regresión lineal se busca clasificar de manera binaria entradas en base a las demás columnas presentes en el dataset.

Transformación de los datos

Al no haber presente datos faltantes en el dataset o datos que no sean válidos para el modelo de regresión, no hizo falta implementar algún tipo de transformación para que pudiera ser utilizable el dataset.

Definición de variables

Para conseguir las variables independientes tome todas las columnas presente en el dataset, y posteriormente elimine la columna de target al ser el resultado, y por lo tanto la variable dependiente, a su vez elimine la columna de id al ser una variable única que no está relacionada con el resultado.

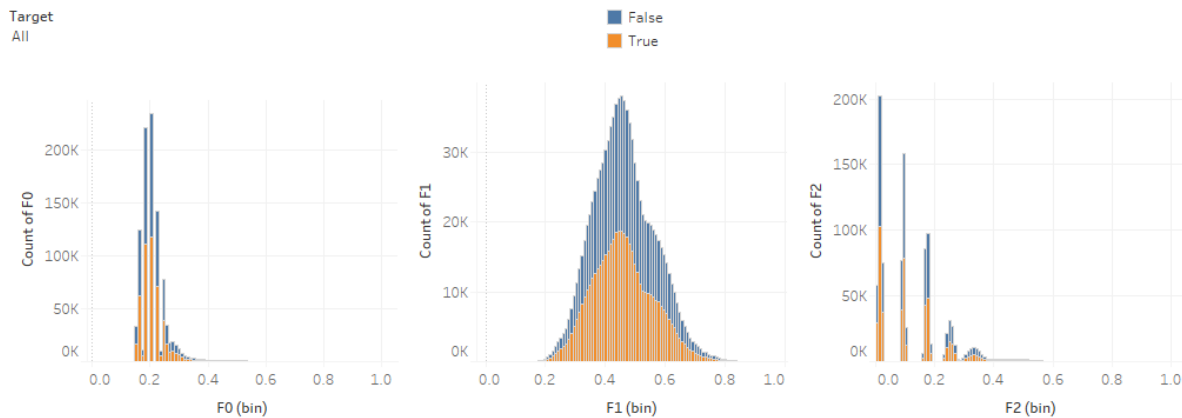
Una vez teniendo las variables independientes las convertí en un vector para poder utilizarlas en el modelo de regresión.

Resultados del modelo

El primer modelo es bueno determinando cuando el resultado va a ser negativo teniendo 80.4% de aciertos en verdaderos positivos, sin embargo su desempeño general es inferior teniendo un 75.6% de precisión.

Ajustando el modelo no vemos una diferencia en el rendimiento, por lo que podemos decir que el modelo llego al límite de las capacidades predictivas, por lo cual si quisiéramos mejorar el rendimiento tendríamos que realizar un procesamiento de los datos o utilizar otro tipo de modelo que nos de mejores resultado con los mismos datos de entrada.

Tableau



Dashboard

La naturaleza de los datos impiden realizar un entendimiento del negocio y un entendimiento de los datos, razón por la que se me dificultó realizar un dashboard en tableau, limitandome a realizar un ejemplo utilizando las primeras variables, con el fin de mostrar cómo se podrían representar el resto de variables en función de si se aplicara un filtro para mostrar los valores que toma según el resultado.

Considerando que los datos no nos dan información con la cual hacer representaciones más complejas, y desconocemos la relación entre los mismos datos, considere que lo mejor para representar los datos sería el mostrar la

distribución de los mismos con un filtro para identificar si cambia la distribución en función del resultado.

Un detalle que agregue fue una imagen, la cual al hacer clic redirige hacia el dataset en kaggle.