

Taller 2

Juan Fernando Guzmán Castaño

1.1. ¿Qué variables son importantes para predecir el valor?

Las variables seleccionadas fueron las siguientes:

- ESTU_GENERO
- ESTU_FECHANACIMIENTO (Error con los dataframe, generando 100% de NAN o muchísimos NAN)
- ESTU_EXTERIOR
- PERIODO
- ESTU_DEPTO_RESIDE
- ESTU_AREARESIDE
- ESTU_TITULO OBTENIDO BACHILLER
- ESTU_VALOR MATRICULA UNIVERSIDAD
- FAMI_ESTRATO VIVIENDA
- FAMI_TIENE INTERNET
- FAMI_TIENE COMPUTADOR
- FAMI_TIENE SERVICIO TV
- FAMI_TIENE CONSOLA VIDEOJUEGOS
- ESTU_HORAS SEMANA TRABAJA
- ESTU_PAGO MATRICULA
- GRUPO REFERENCIA
- ESTU_INSE_INDIVIDUAL
- ESTU_NSE_IES

1.2. ¿Existen nulos?, ¿cómo se deben imputar?

Si, dentro de cumpleaños o como las dos ultimas variables, estu_inse_individual y estu_nse_ies, que contenían nulos.

Estos nulos se imputan con la media, moda o mediana, dependiendo del sesgo que exista se puede usar una o otra, en mi caso use la media.

2.3. ¿Cuál es el mejor R squared?Cuál es el MAPE y el MSE.

R2: 0.24888130937775754
MAPE: 21375464987832.48
MSE: 414.03471036915107

Estos son los valores para la regresión.

2. Crear un modelo de KNN

Los Resultado obtenidos por el knn son.

```
K = 5
MSE en entrenamiento: 311.2710780631017
MAPE en entrenamiento: 6351148818742.868
MSE en prueba: 469.0232097879087
MAPE en prueba: 21232082960112.305
-----
K = 10
MSE en entrenamiento: 353.05281868181174
MAPE en entrenamiento: 7390427716355.33
MSE en prueba: 434.78564240629396
MAPE en prueba: 20140840117619.23
-----
K = 20
MSE en entrenamiento: 378.143037012346
MAPE en entrenamiento: 7725306472252.676
MSE en prueba: 419.81226412695565
MAPE en prueba: 19763620369596.93
-----
K = 30
MSE en entrenamiento: 388.18528239666443
MAPE en entrenamiento: 7902368802957.02
MSE en prueba: 416.9774452986595
MAPE en prueba: 19646861876161.46
```

Según los resultados, comparando a la regresión y al modelo knn, en training resulta ser mucho mejor la regresión, debido a que tiene un mse mucho menor que todos los knn entrenados desde 5, 10, 20 y 30.

3 crear un modelo de GBM.

Los resultados para el GBM es:

```
MSE en entrenamiento: 408.89303475968336
MAPE en entrenamiento: 7975211347567.93
MSE en prueba: 414.2159863972925
MAPE en prueba: 21276385326141.668
```