

Diamonds – datamad0120

<https://www.kaggle.com/c/diamonds-datamad0120/leaderboard>

<https://github.com/sowe01/Proyecto-ML.git>

Utilizo :

Me centro en scikit-learn, son unas herramientas para predecir analisis de datos, para ello importo:

RandomForestRegressor
DecisionTreeRegressor
enable_hist_gradient_boosting
HistGradientBoostingRegressor
GradientBoostingRegressor
KNeighborsRegressor
LinearSVR
DecisionTreeClassifier
LogisticRegression

Reto:

Hay que predecir el precio de un diamante segun sus características.
Tenemos 2 datasets:

-diamonds_train.csv
-diamonds_test.csv

diamonds_train con lo precios y
diamonds_test sin ellos
para que hagamos la prediccion

Pá Diamante el !



Proceso:

limpiamos diamonds_train, y test, esta es la parte mas importante

Compruebo las columnas tipo objeto y les asigo valores relacionados entre si para tener la columna tipo numerica

luego paso a entrenar un primer modelo prediciendo los resultados de test en la columna price, y lo subo a kaggle para ver los resultados

Una vez hecho esto entreno con mas modelos

Conclusion:

El modelo histboosting3 jjes el ganador provisional!!!

#	Team Name	Notebook	Team Members	Score	Entries	Last
1	Os voy ganando !!!			492.48166	9	44m
Your Best Entry ↑ Your submission scored 512.86478, which is not an improvement of your best score. Keep trying!						
2	vagabundoo			494.70085	9	8h
3	Patricia Carmona Pulido			504.85127	15	1d
4	Xabier García de Vicuña Zuaz...			506.22513	8	2d
5	Amanda Gálvez Corell			507.15462	15	1h
6	Cristina Camargo			508.98554	7	15m
7	Maria Roig			509.57140	5	2h
8	Marlena Anna Polak			510.79515	4	1h