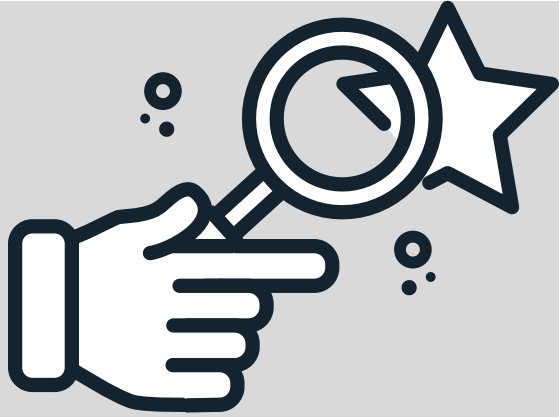


Diamond Competition

¿CÓMO PREDECIR EL PRECIO DE UN DIAMANTE?

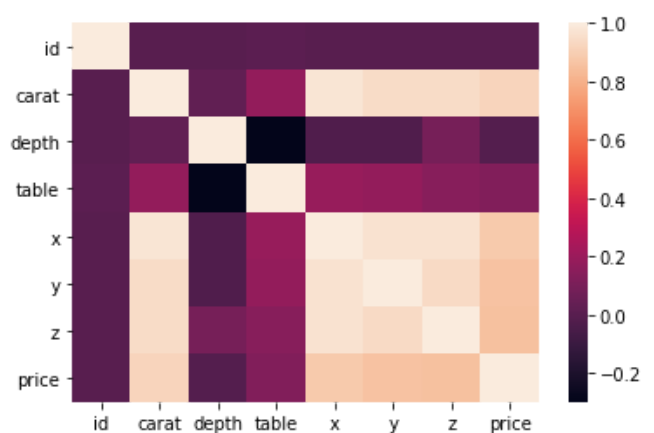


OBJETIVO

Desarrollo de multiples modelos de predicción para un problema de regresión. Utilizando el Dataframe de kaggle de Diamantes.

ANÁLISIS DE DATOS

Una vez limpios los datos, se busca la correlacion entre ellos y se observa una correlacion fuerte entre el Precio Vs. x,y,z ,carat



MÉTODOS PARA ANALISIS DE REGRESIÓN



DecisionTreeRegressor

ExtraTreesRegressor

GradientBoostingRegressor

HistGradientBoostingRegressor

RandomForest

MÉTRICAS UTILIZADAS

R^2

MSE

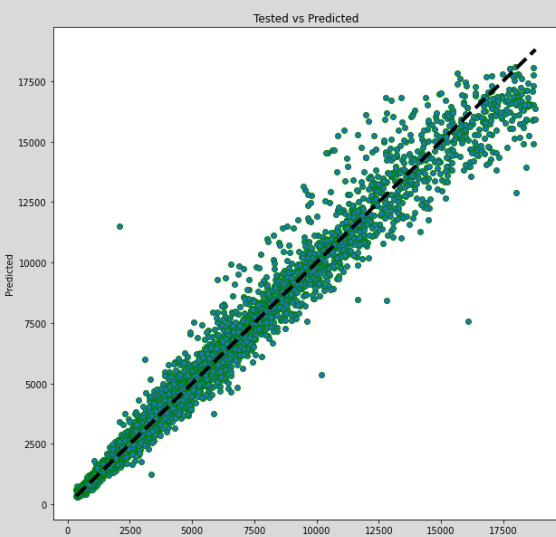
MAE

Explained variance score

```
R2_score 0.9795876891565237
Mean squared error 581.8007314368067
Mean absolute error 303.3389528566118
explained variance score 0.9795953219248649
```

```
: scores = cross_val_score(model,X,y, cv=10)
print(np.mean(scores))
```

```
0.9800925585775719
```



CONCLUSIÓN

De los modelos probados el que tiene mejor MSE y se ajusta mas a la realidad de la predicción es el Hist Gradient Boosting Regressor y un cross validation de 0,98

FUENTE/S Y AYUDAS:

<https://scikit-learn.org/stable/>

<https://colab.research.google.com/>

<https://machinelearningmastery.com/>

[metrics-evaluate-machine-learning-algorithms-python/](#)



By: Joyce Acevedo

Data Agosto 2020