# Encode variables with more than 2 Classes

df = pd.get\_dummies(df, columns= [i for i in df.columns if df[i].dtypes=='object'],drop\_first=True)

(Linear Regression)Baseline

Baseline Model's score: 0.91 (avg) 0.01 (std)

Wall time: 8.53 s

KNN Mean score: 0.94 (avg) 0.00 (std)

Wall time: 24.1 s

MLPReg Mean score: 0.95 (avg) 0.02 (std)

Wall time: 1min 12s

DT Mean score: 0.95 (avg) 0.00 (std)

Wall time: 975 ms

RF Mean score: 0.97 (avg) 0.00 (std)

Wall time: 37.9 s

Questions:

1. Remove features?
2. Remove outliers

RandomForestRegressor(n\_estimators=500, random\_state=42)

{'max\_depth': None, 'min\_samples\_split': 2, 'n\_estimators': 500}

0.9738123036207386

Wall time: 36min 23s

Final Model (RandomForestRegressor) Metrics -

R2 Score: 0.9766406237017738

Mean Absolute Error: 289.91854664685485

Mean Square Error: 364291.8654780507

StandardScaler =>

array([[-0.94382039, -1.0816645 , 1.14451719, ..., 1.84365604,

-0.27053165, -0.32175912],

[-1.04902901, 0.10942298, -0.2088355 , ..., -0.54240052,

-0.27053165, 3.10791503],

[-1.02798729, 0.45974282, -1.11107062, ..., -0.54240052,

-0.27053165, -0.32175912],

...,

[-0.83861176, 0.17948694, -0.65995306, ..., -0.54240052,

-0.27053165, 3.10791503],

[ 0.2134745 , 0.7399987 , 0.69339963, ..., -0.54240052,

-0.27053165, -0.32175912],

[ 0.86576798, 0.81006266, 0.24228207, ..., -0.54240052,

-0.27053165, -0.32175912]])

User input:

Carat: 1

Cut: Fine

Color: F

Clarity: I1

X: 8

Y: 42

Z: 28

Depth: 55

Table: 72

StandardScaler??? Then pass the model for prediction?