Assignment – 2
Total Points = 80

Housing Price Prediction Problem in Kaggle

https://www.kaggle.com/c/house-prices-advanced-regression-techniques/overview

Description of Data and Submission Format

https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data

Task Description

1. Perform following steps on data (if you omit any step please specify why you did it):

a. Missing data (Null Value) and Categorical data Handle – 10 points

b. Find correlation matrix (in terms of target variable/feature) – 10 points

c. Perform proper feature scaling to transform feature columns and target variable (if necessary) - 10 points

2. You have to use supervised learning models for this problem and report the output you get for each of the following method:

a. *Simple Linear Regression with Regularization (use both L1 and L2 regularization) - 20 Points*

b. *Decision Tree (with maximum depth restriction and no restriction) – 10 Points*

c. *Random Forest (Try with various estimators count) – 10 points*

3. Perform PCA and then apply decision tree algorithm, in your opinion does that improve the results compared to regularization tasks done in step 1(a)? - 10 points

NOTE: This is regression problem, so linear regression, decision tree and random forest all will use the regression version: DecisionTreeRegressor, RandomForestRegressor etc.

SAMPLE SOLUTIONS

1. https://www.kaggle.com/iavinas/house-pricing-simple-solution-top-30

2. https://www.dataquest.io/blog/kaggle-getting-started/

3. https://www.kaggle.com/vjgupta/reach-top-10-with-simple-model-on-housing-prices

Google search to find easier solution. Every answer can be found in the internet. I just want you to compile them and give comments on everything you do as part of this code with little one line explanation why you are doing that.