**Machine Learning Assignment**

**Name:Salma Ayman**

**Id:43-8182**

* Importing libraries such as pandas, numpy ,and.sklearn

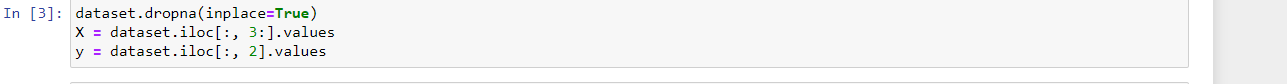
Graphical user interface, application

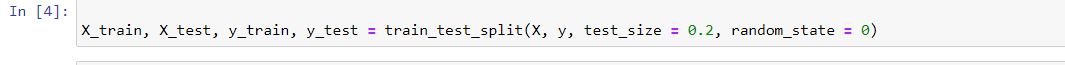
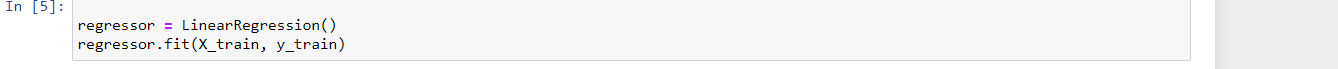
Description automatically generated

* Importing the dataset



* The X is independent variable array and it represents the data set features and y is the dependent variable vector and it represents the price . Note the difference between the array and vector. The dependent variable must be in vector and independent variable must be an array itself

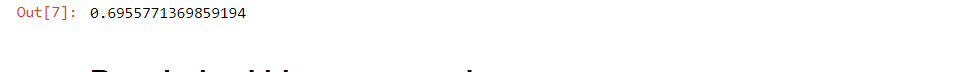


* We need to split our dataset into the test and train set where in training I will train my model and in the testing I will test my model accuracy.
* We need to fit X\_train (training data of matrix of features) into the target values y\_train. ] and train the x train data using the y train data that I have split using linear regression
* Graphical user interface, text

  Description automatically generated with medium confidenceFitting polynomial regression to the dataset
* predict method makes the predictions for the test set. the input is the test set. The parameter for predict must be an array because input is X\_test
* Graphical user interface, text, application, email

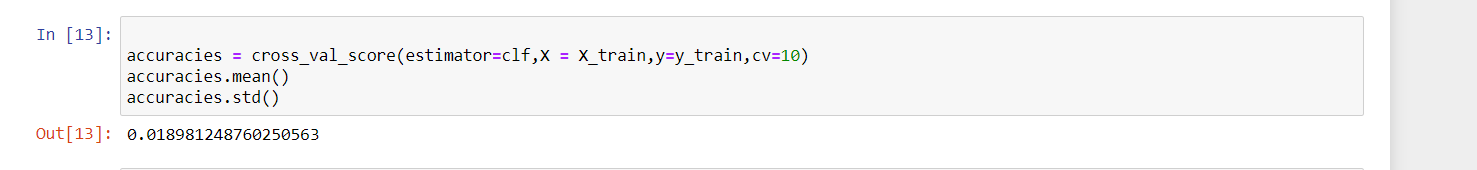
  Description automatically generatedfrom predicting the y values using the x test values I’m giving my model accuracy score.



* The output:
* Graphical user interface, text, application

  Description automatically generatedfirst training the x train data using the y train data that I have split using regularized linear regression .then trying to predict trying to predict the y values from the x test values using regularized linear regression .
* Text

  Description automatically generatedK-folds Allows , to split data into train and test sets using train/test indices. That split the dataset into k consecutive folds so I’m splitting my data to 5 sections then I’m dividing my data again using this 5 different sections ,then I’m training my data
* using the new sections and calculating the accuracy score for these sections.
* Graphical user interface, text, application

  Description automatically generatedImplementing the KFold cross validation using regularized linear regression.
* Calculating the accuracies for all the folds.
* Text

  Description automatically generatedImplementing StratifiedKFold that shuffles the data, after that splits the data into parts .and it will use each part as a test set.
* it only shuffles data one time before splitting.

Text

Description automatically generated

* Implementing the StratifiedKFold using the regularized linear regression.