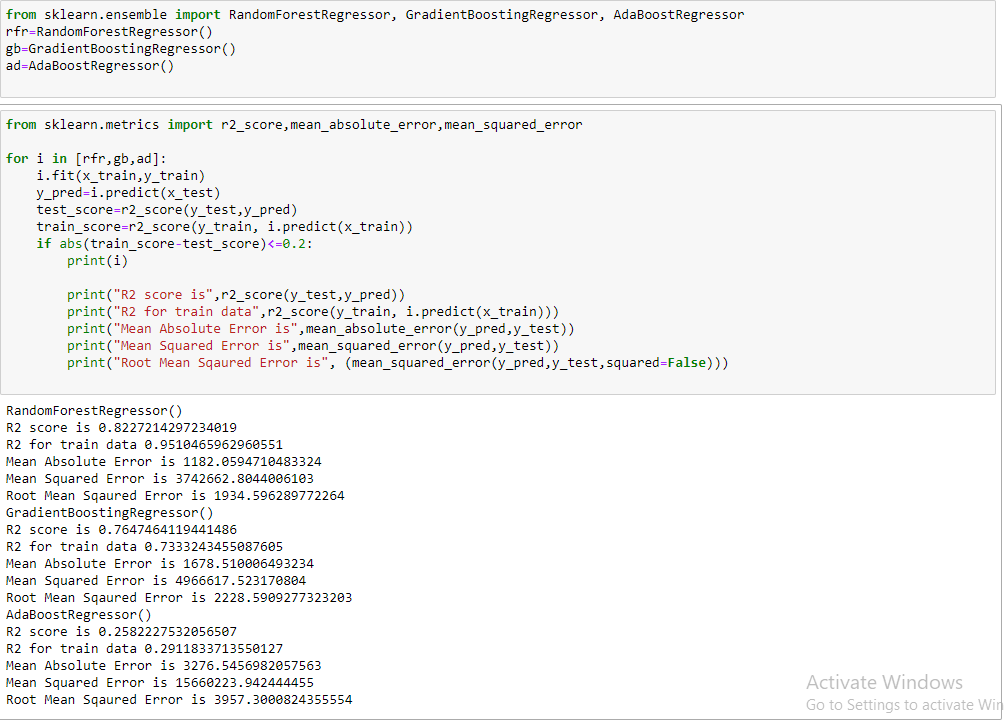
**Model Building**

Now our data is cleaned and it’s time to build the model. We can train our data on different algorithms. for this project we are applying four regression algorithms. The best model is saved based on its performance.

**Using Ensemble Techniques**

**RandomForestRegressor, GradientBoostingRegressor, AdaBoostRegressor**

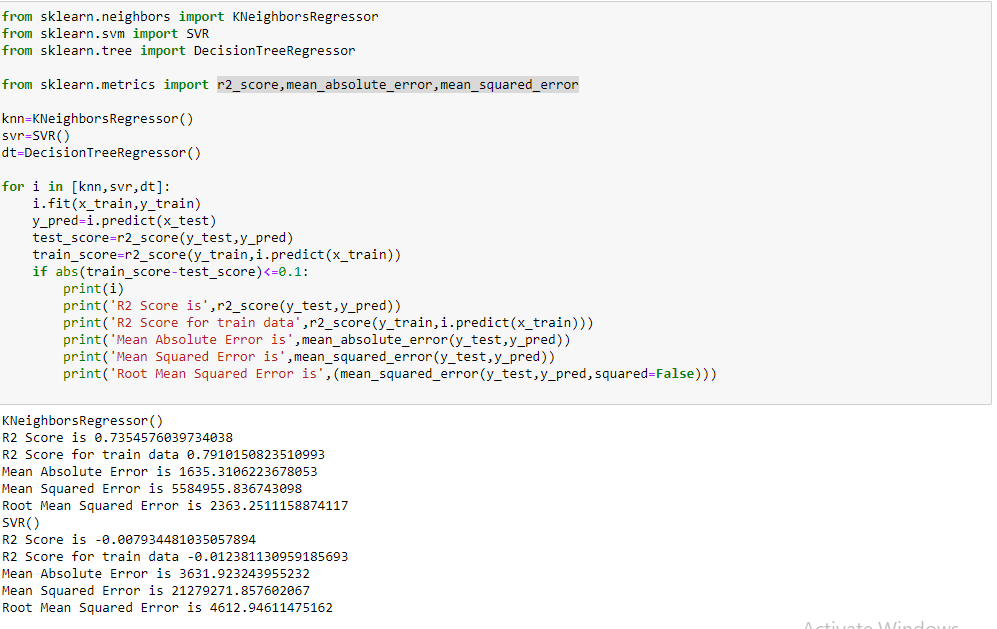
A function named RandomForest, GradientBoosting, AdaBoost is created and train and test data are passed as the parameters. Inside the function, RandomForest, GradientBoosting, AdaBoost algorithm is initialized and training data is passed to the model with .fit() function. Test data is predicted with .predict() function and saved in new variable. For evaluating the model, r2\_score, mean\_absolute\_error, and mean\_squared\_error report is done.



**Regression Model**

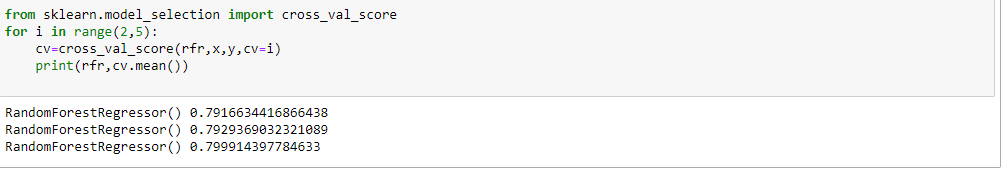
**KNeighborsRegressor, SVR, DecisionTreeRegressor**

A function named KNN, SVR, DecisionTree is created and train and test data are passed as the parameters. Inside the function, KNN, SVR, DecisionTree algorithm is initialized and training data is passed to the model with .fit() function. Test data is predicted with .predict() function and saved in new variable. For evaluating the model, r2\_score, mean\_absolute\_error, and mean\_squared\_error is done.



### Checking Cross Validation For RandomForestRegressor

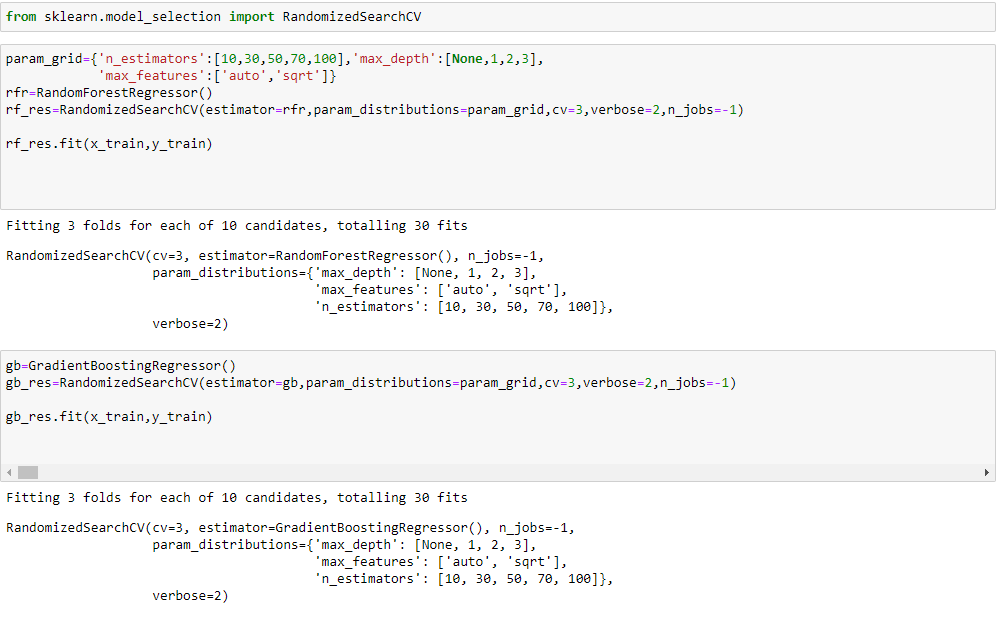
We perform the cross validation of our model to check if the model has any overfitting issue, by checking the ability of the model to make predictions on new data, using k-folds. We test the cross validation for Random forest and Gradient Boosting Regressor.



**Hypertuning The Model**

RandomSearch CV is a technique used to validate the model with different parameter combinations, by creating a random of parameters and trying all the combinations to compare which combination gave the best results. We apply random search on our model.

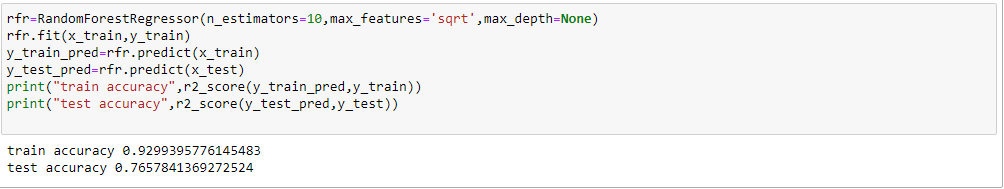
From sklearn, cross\_val\_score is used to evaluate the score of the model. On the parameters, we have given rf (model name), x, y, cv (as 3 folds). Our model is performing well.



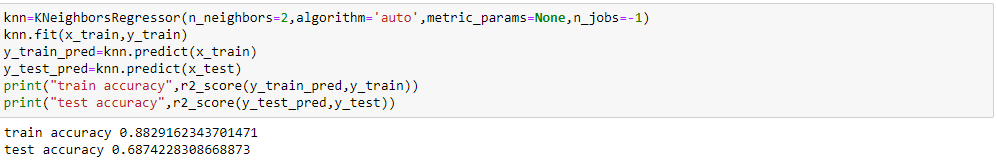
Now let’s see the performance of all the models and save the best model

### Accuracy

Checking Train and Test Accuracy by RandomSearchCV using RandomForestRegression Model



Checking Train and Test Accuracy by RandomSearchCV using KNN  Model2



By Observing two models train and test accuracy we are getting good accuracy in RandomForestRegression

### Evaluating Performance Of The Model And Saving The Model

From sklearn, cross\_val\_score is used to evaluate the score of the model. On the parameters, we have given rfr (model name), x, y, cv (as 3 folds). Our model is performing well. So, we are saving the model by pickle.dump().

Note: To understand cross validation, refer this link. <https://towardsdatascience.com/cross-validation-explained-evaluating-estimator-performance-e51e5430ff85>.

