**California House Price Prediction:-**

In this Project I have built a website that shows you little information of the project and then we have predict feature where you can give your input values accordingly it predicts the price and gives output for which we have used California housing dataset from Kaggle. The website is made using fundamentals of html, CSS, bootstrap, jQuery, and machine learning.

**Technologies:**

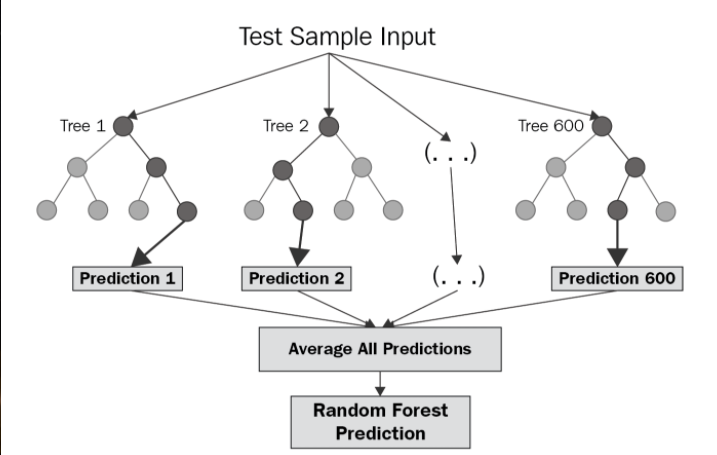
* Python as Programming Language
* Pandas for Data Cleaning
* matplotlib for Data Visualization
* sklearn for Model Building
* Flask for Integration
* HTML, CSS , Bootstrap & jQuery for Frontend

For detailed information refer to requirements.txt file.

**Machine Learning Model:**

**Random Forest Regression Model:-**

Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression. ... A Random Forest operates by constructing several decision trees during training time and outputting the mean of the classes as the prediction of all the trees

****

**The diagram above shows the structure of a Random Forest. You can notice that the trees run in parallel with no interaction amongst them. A Random Forest operates by constructing several decision trees during training time and outputting the mean of the classes as the prediction of all the trees. To get a better understanding of the Random Forest algorithm, let’s walk through the steps:**

**1.Pick at random** k **data points from the training set.**

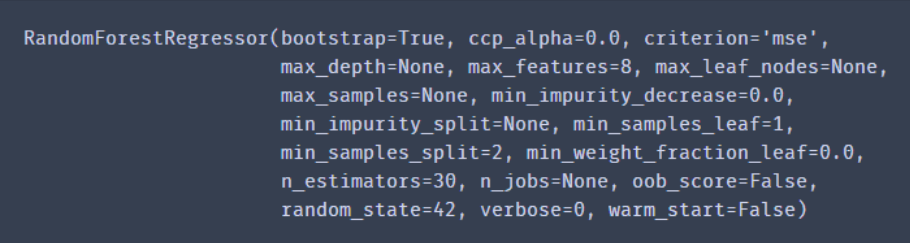
2.Build a decision tree associated to these k data points.

3.Choose the number N of trees you want to build and repeat steps 1 and 2.

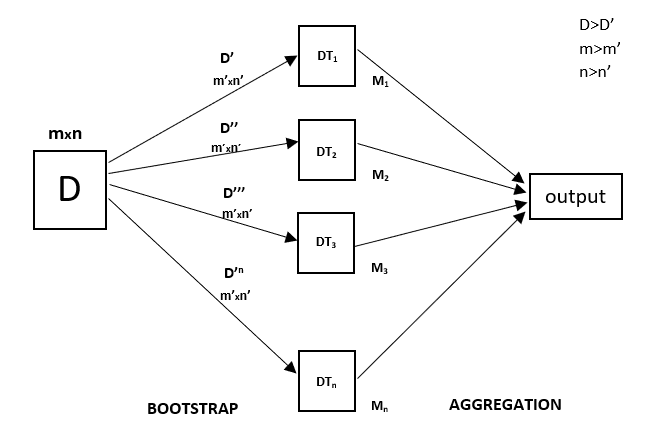
4.For a new data point, make each one of your N-tree trees predict the value of y for the data point in question and assign the new data point to the average across all of the predicted y values.

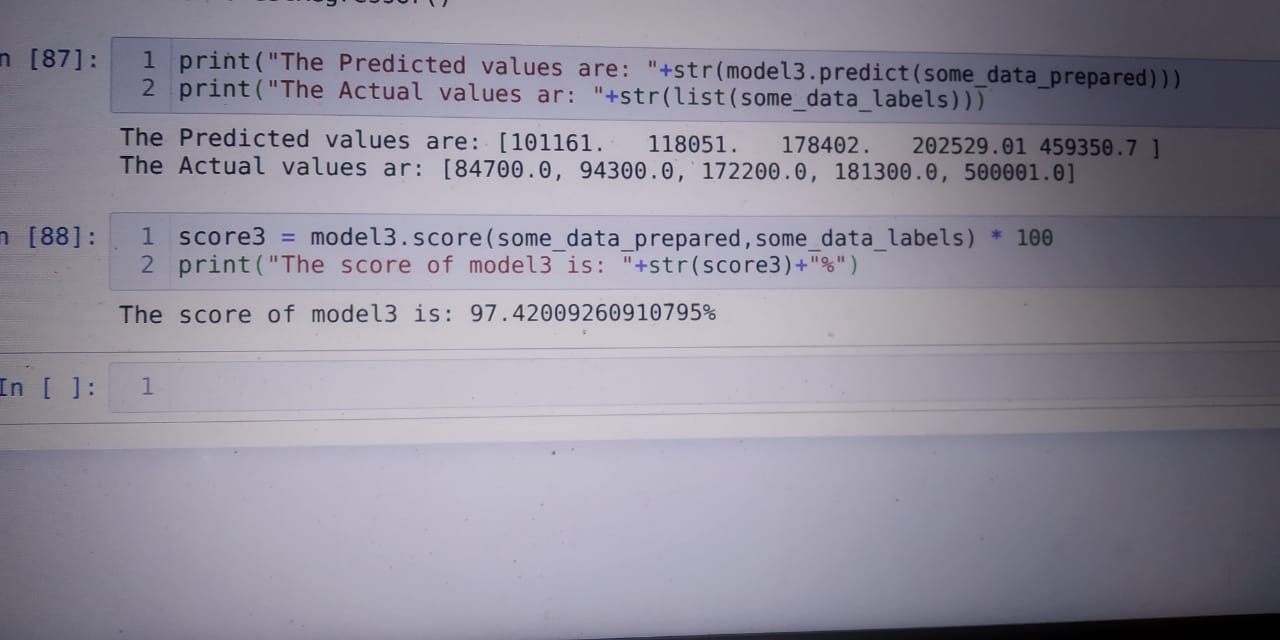
A Random Forest Regression model is powerful and accurate. It usually performs great on many problems, including features with non-linear relationships. Disadvantages, however, include the following: there is no interpretability, overfitting may easily occur, we must choose the number of trees to include in the model.

**This is available in sklearn package in python.**

****

**The bootstraping and agregation in random forest regression is shown below in th image:-**

**And the score of the RandomForestRegreesor model on California housing prices dataset is as shown blow in the image:-**

****