**A PROJECT REPORT**

**ON**

**REAL ESTATE PRICE PREDICTION SYSTEM**

**USING ML REGRESSION ALGORITHIM**

Submitted in partial fulfilment for the requirement of the award of

TRAINING

IN

Data Analytics, Machine Learning and AI using Python



Submitted By

**Ashutosh Tiwari** (SRMS college of engineering technology & Research, Bareilly)

Under the guidance of

**Mr. Bipul Shahi**

**ACKNOWLEDGEMENT**

My sincere gratitude and thanks towards my project paper guide Bipul Shahi, Corporate Trainer, Developer, Traveller !!! IOT, Artificial Intelligence, Robotics, Cloud Computing, Android Apps !!!

It was only with his backing and support that I could complete the report. He provided me all sorts of help and corrected me if ever seemed to make mistakes. I have no such words to express my gratitude. And at last but not the least, I acknowledge my dearest parents for being such a nice source of encouragement and moral support that helped me tremendously in this aspect. I also declare to the best of my knowledge and belief that the Project Work has not been submitted anywhere else.

**INTRODUCTION**

The goal of that project is to predict the efficient ,reliable price of the house on the basis of the current and future market trends. People looking to buy a new home and the wants that home to be more conservative with their budgets and market strategies. This price prediction system predict the approx. nearar price of the house. on this basis of the previous datasets of the price of the houses having different features. This model takes the house feature detail from the customer’s and on the basis of the trained model predict the approximate price of that house. This application will help customers to invest in an real estate without approaching an agent. It also decreases the risk involved in the transaction.

**Problem statement**

Currently the problem of the house buyers is that they don’t predict the approximate price of any house . because every buyer does not comes from the real estate service background.so for the prediction they approaching to an agent.and the agent fee added to the price of house. And the house price increases.but with the help of this project one can predict the efficient ,reliable price of the house on the basis of the current and future market trends. andthis application will help customers to invest in an real estate without approaching an agent. It also decreases the risk involved in the transaction.

**Technology and Concepts**

**Machine Learning**

Artificial intelligence (AI) has so many application like machine learning ,deep learning ,neural network etc. Machine learning is one of them that provides systems the ability to automatically learn and improve from experience or previous database without being explicitly programmed. **Machine learning focuses on the development of computer**

**programs** that can access data and use it learn for themselves and gives some predictions after analysing the data.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. **The primary aim is to allow the computers learn automatically** without human intervention or assistance and adjust actions accordingly.

Machine learning has so many different type of algorithms for different problems like classification ,regression ,clustering.

**Type of machine learning algorithms**

* **Supervised machine learning algorithms** can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

There are two types of supervised learning

1.Classification

2.Regression

* **Unsupervised machine learning algorithms**are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn’t figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

There are two types of supervised learning

1.Clustering

2.Association

* **Reinforcement machine learning algorithms**is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

Here for this project I used the Regression algorithms of Supervised machine learning. Regression is a process of finding the correlation between dependent and independent variable.it helps in predicting the continuous variable such as real estate house price prediction etc.

Here I used different types of regression algorithms

1.multiple linear regression

2.decision tree regression

3.random forest regression

4.support vector regression

5.KNeighbour’s algo

**Package and module used**

1.Python 3

2.Scikit learn

3.Pandas

4.Numpy

5.Matplotlib

**Requirements**

1.Python 3 interpreter (or anaconda navigator)

2.jupyter notebook (or PyCharm, spider etc)

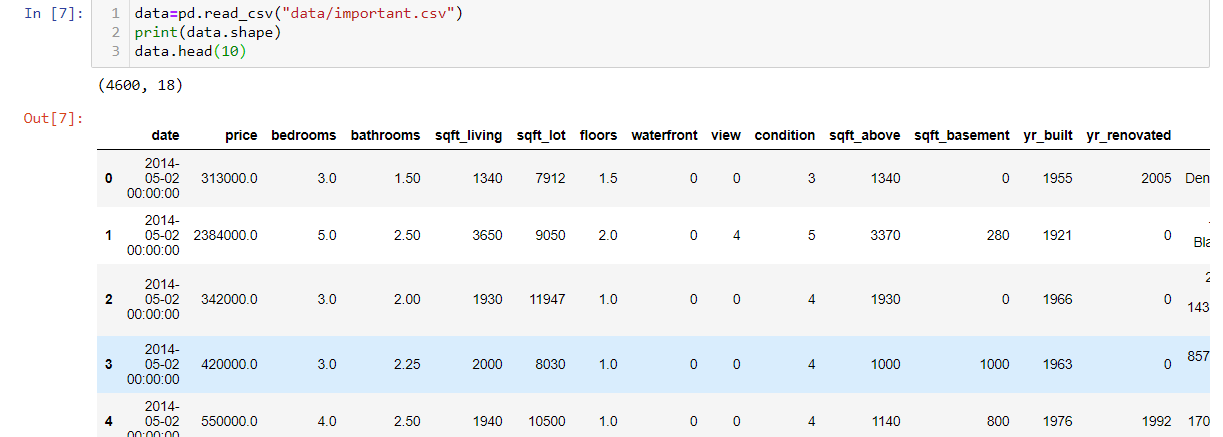
2.Ram (4gb min)

3.Processor (i5 min)

**Description of the dataset**

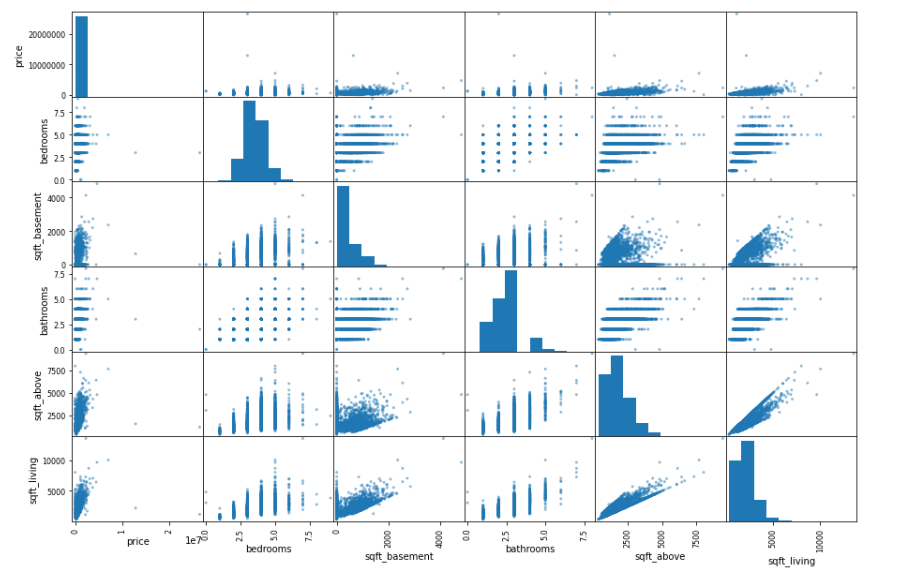
**dataset**

Here I used the dataset of the USA real estate in which the different types of attribute information exist and also price is there.



**Scatter matrix of attributes and label**

****

****

**Experimental setup**

**Model Building**

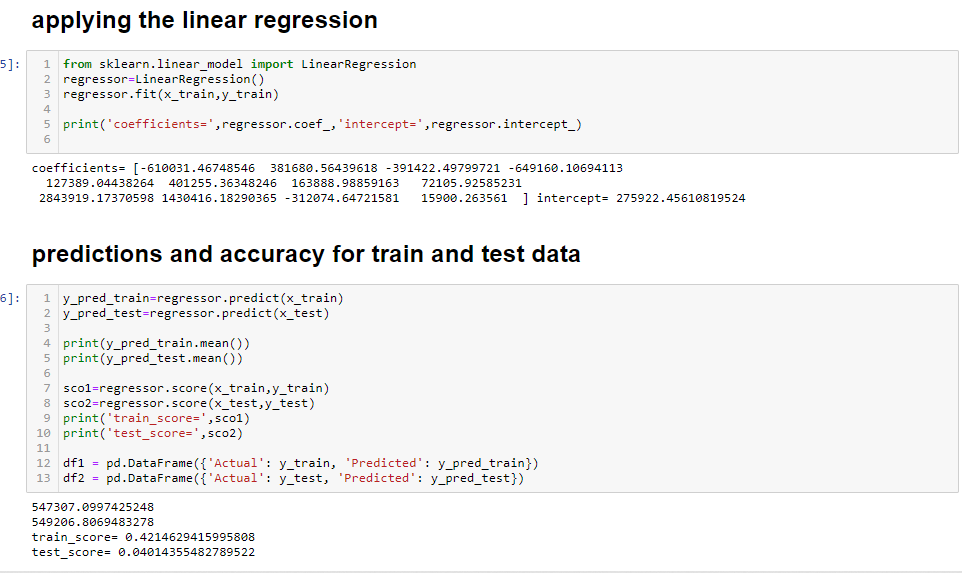
**1.Model building using multiple linear regression**

Multiple regression generally explains the relationship between multiple independent or predictor variables and one dependent or criterion variable.  A dependent variable is modeled as a function of several independent variables with corresponding coefficients, along with the constant term.  Multiple regression requires two or more predictor variables, and this is why it is called multiple regression.

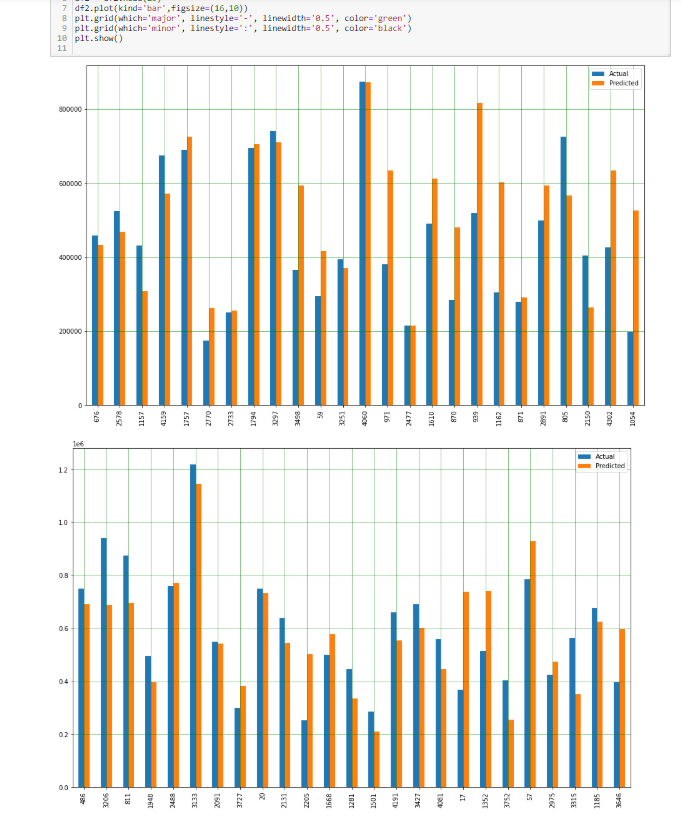
The multiple regression equation explained above takes the following form:

y = b1x1 + b2x2 + … + bnxn + c.

Here, bi’s (i=1,2…n) are the regression coefficients, which represent the value at which the criterion variable changes when the predictor variable changes.

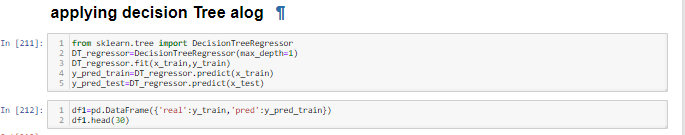


**Graphical representation of the predictions for train and test data**

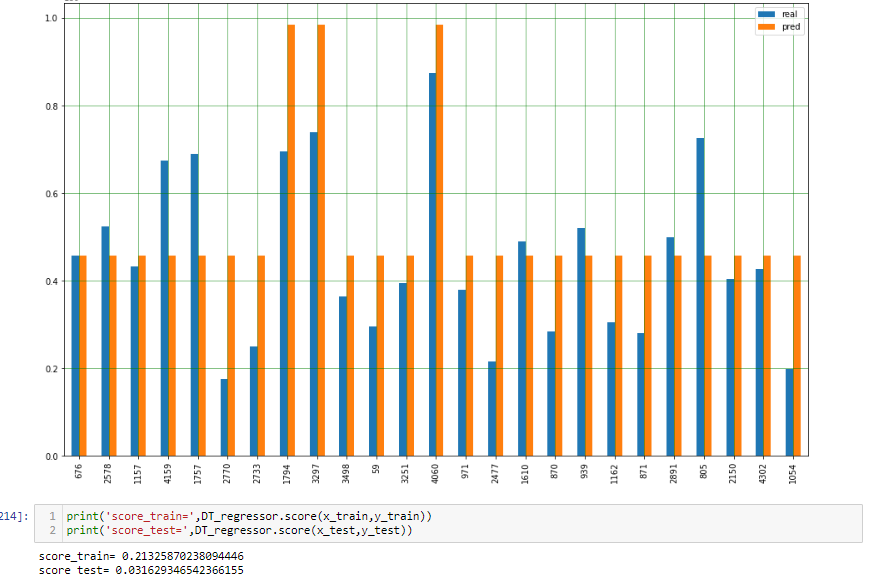
****

**2.Model building using decision tree regression**

Decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.

****

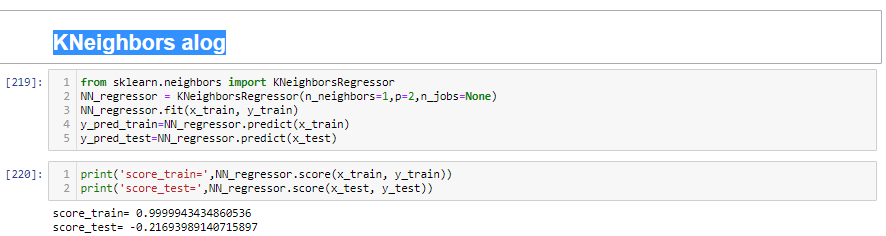
**Graphical representation of the predictions for train and test data**

****

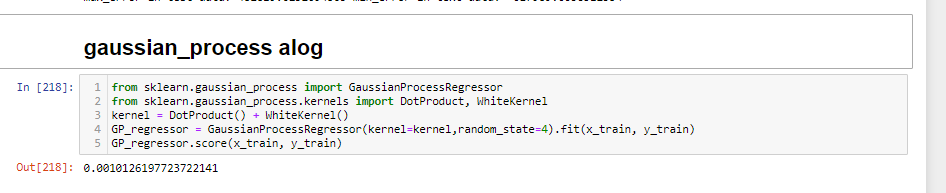
****

**3.Model building using KNeighbour’s algo**

Regression based on k-nearest neighbour’s. Regression with scalar, multivariate or functional response. The target is predicted by local interpolation of the targets associated of the nearest neighbour’s in the training set.

****

**4.Model building using guassian\_process algo**

****

**Conclusions**

This paper has practiced so many different types of machine learning technique and different models for data training attempting to predict the nearer prices of the houses to its original prices in the current scenario. Thus, this study settled on regression a given data of houses and train different regression model using five different algorithms and consequently testing its accuracy. Here I used the multiple regression algo, decision tree algo etc and also tested the score of each model. Decision tree perform well. But it gives 81% accuracy on the train data set but does not working well on the test dataset. I think this is because of the dataset problem.

**Bibliography**

* <http://www.sthda.com/>
* <https://www.kaggle.com>
* <https://github.com/>
* GitHub link of project-<https://github.com/ashutoshtiwari-jnv/ML_project>