HOUSE PRICE PREDICTION

V. Vijay Bhasker Reddy
Associate Professor,
CSE, Sreyas institute of engineering and Technology, Telangana, India
vijayreddy@sreyas.ac.in
Burugu Chandana
CSE, Sreyas institute of engineering and Technology, Telangana, India
Buruguchandana27@gmail.com
Reddy Sindhu Reddy
CSE, Sreyas institute of engineering and Technology, Telangana, India
Sindhureddy6602@gmail.com
Tandava Charitha
CSE, Sreyas institute of engineering and Technology, Telangana, India
tandavacharitha557@gmail.com
Sravya Siripragada
CSE, Sreyas institute of engineering and Technology, Telangana, India
Sravyasrihi1329@gmail.com

Abstract

Because house prices rise every year, a mechanism to forecast future house values is required. House price prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house. Physical conditions, concept, and location are only a few of the aspects that determine the price of a home. Usually, House price index represents the summarized price changes of residential housing. While for a single-family house price prediction, it needs a more accurate method based on location, house type, size, build year, local amenities, and some other factors which could affect house demand and supply. A practical and composite data pre-processing, creative feature engineering method is investigated with limited dataset and data features. This model is used to predict the house prices so as to cut down the complications faced by the customers. The present method where customers reach real-estate agents and search for houses in their budget, and should analyze whether a particular price is accurate or not. To overcome this our proposal is used. This system makes optimal use of the Machine Learning Algorithms. By extracting data from datasets of different houses, preprocessing the data and model is built using that data using Regression. The algorithm used for the model building is KNN (K Nearest Neighbor) Algorithm. This system design is modularized into various categories.

I. INTRODUCTION:

A practical and composite data pre-processing, creative feature engineering method is investigated with limited dataset and data features. Data is used to train predictive models, which results in reasonably accurate outcomes. Without data, we can't train the model. Machine learning involves building these models from data and using them to predict new data. Every firm in the real estate industry today is working hard to gain a competitive advantage over its competitors. There is a need to simplify the process for normal human beings while providing the best results. Housing price trends are a source of concern for both buyers and sellers since they reflect the current economic climate. Manually predicting housing values is a difficult undertaking that is rarely correct. So that, we need make a model that can give us a good house price prediction based on house features.

EXISTING SYSTEM:

The present method is that the customer approaches a real estate agent to manage his/her investments and suggest suitable estates for his investments. Many people are facing problems in buying suitable properties on their own. A website named 99acers exists.com which gives suggestions based on customer requirements and budget range. Sometimes sellers can't identify whether they are dealing with the right agents at the right price. And there are some existing models which predict using different algorithms like Linear regression, Lasso regression and Ridge regression which are not accurate compared to our mode

Disadvantages:

There are many disadvantages in existing system like:

- There is no prediction of price in 99acres website, it only recommends the houses based on budget as well as well as characteristics.
- From predictive models we can say that accuracy is low and prediction may go wrong sometimes.

II. PROPOSED METHOD

Nowadays, e-education and e-learning is highly influenced. Everything is moving away from manual processes and toward automated ones. The goal of this project is to forecast house prices in order to minimise the customer's issues. The current strategy entails the customer approaching a real estate agent to handle his or her investments and recommend suitable estates. However, this strategy is hazardous because the agent may estimate incorrect estates, resulting in the customer's capital being lost. The manual approach now in use on the market is obsolete and fraught with danger. There is a need for an updated and automated system to address this flaw. Data mining algorithms can be used to assist investors in selecting an acceptable property based on their stated needs.

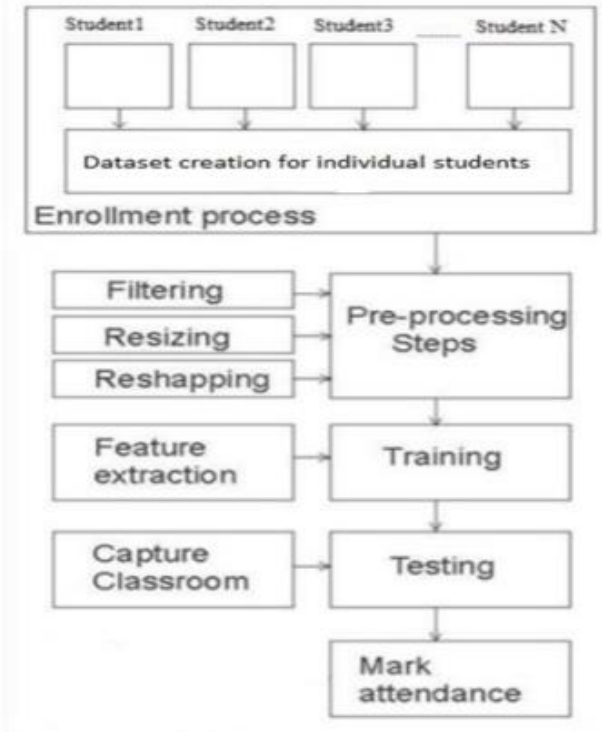
Advantages of proposed system

- Predictive model that can accurately predict house price.
- Analyze the house features accurately so that we can avoid buying houses at inaccurate prices given by owners.

- More effective on large datasets compared to other algorithms

A. PROPOSED ARCHITECTURE

The relationship between different components might be depicted using a system architecture diagram. Typically, they are made for systems that comprise both hardware and software, which are shown in the diagram to explain how they interact.



B. IMPLEMENTATION

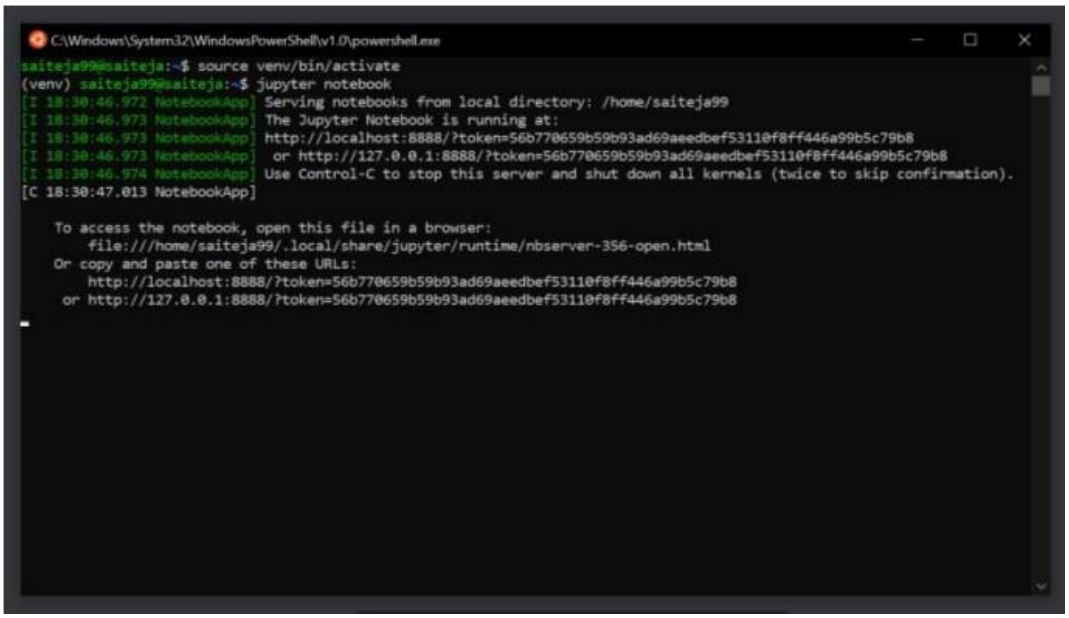
Different functions are implemented using different python libraries in our project to build predictive models. Main libraries used for our project are pandas, Matplotlib, NumPy, Turi Create etc. For Data preprocessing pandas is used. Different functions like read_csv (), isNull() and drop() are used in the data cleaning process. Matplotlib is used for visualizing data. Different graphs like scatter plots, bar graphs, Histograms, Boxplots are used to understand and examine the data. NumPy library is used to make complex mathematical operations easy. NumPy plays a key role while calculating distance between two houses. Here Euclidean distance is used to find the distance between query houses and houses of datasets. So we get k nearest neighbors using function argsort () which returns indices of k nearest neighbors. Turicreate library is used for building models and handling data in our study, turi create is easy and flexible to use. Seaborn library is used in data preprocessing.

Steps:

- First step is to open ubuntu LTS terminal and move to virtual environment in it. This is done by running command -> source venv/bin/activate Where, venv is the name of the virtual

environment created

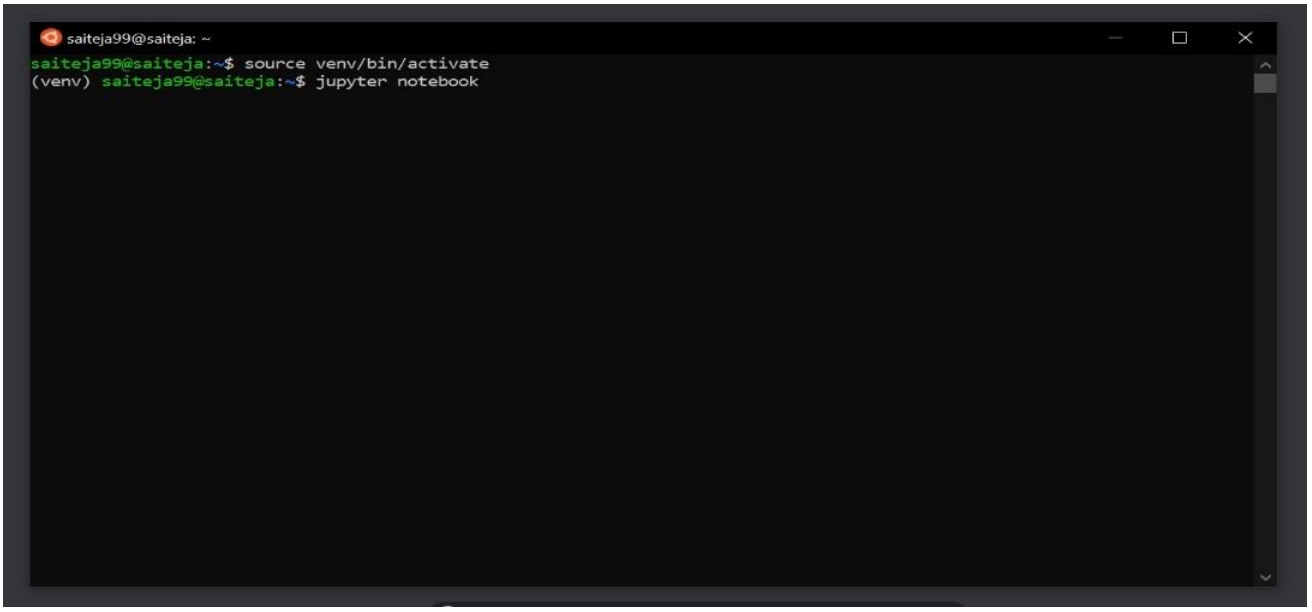
- Now run command -> jupyter notebook after entering into virtual environment.



Screen 1: opening jupyter notebook

Now, jupyter notebook was opened. The data which is to be used is to be uploaded in the jupyter notebook by creating a folder for it. The dataset is accessed in the jupyter notebook through this folder. Create a folder name venv where implementation is done. 24 In this folder create the folder of dataset and also ipynb file in which model building is programmed

- After this, programming was done to obtain the best model. After getting the perfect model, this was run on different test cases. So, for building a model using KNN regression the setup is
- simple and secure. Just a jupyter notebook is enough along with some libraries and also a dataset.



III. RESULT

- Price change for different physical features

```
TestCase1 = {'bedrooms':[4],
             'bathrooms':[4],
             'sqft_living':[1500],
             'sqft_lot':[2250],
             'floors':[1],
             'zipcode':[98136],
             'grade':[8],
             'view':[4],
             'sqft_above':[1500],
             'sqft_basement':[125],
             'years':[25],
             'yr_renovated':[2010]}

TestCase1=get_numpy_testdata(tc.SFrame(TestCase1),feature_list)/norms
TestCase1
AvgHousePrice(best_k,features_train,output_train,TestCase1)
```

```
628327.7777777778
```

```
TestCase2 = {'bedrooms':[2],
             'bathrooms':[3],
             'sqft_living':[1200],
             'sqft_lot':[1550],
             'floors':[2],
             'zipcode':[98136],
             'grade':[5],
             'view':[1],
             'sqft_above':[500],
             'sqft_basement':[125],
             'years':[25],
             'yr_renovated':[2010]}

tc.SFrame(TestCase2)
TestCase2=get_numpy_testdata(tc.SFrame(TestCase2),feature_list)/norms
print(AvgHousePrice(best_k,features_train,output_train,TestCase2))
```

Building 1-NN to 16-NN models and choosing best model

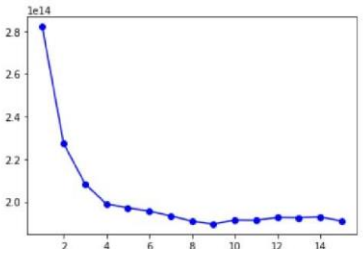
In [143]:
```
rss_all=[]
for k in range(1,16):
    knnval=MultipleHousesPrice(k,features_train,output_train,features_valid)
    knnval=sum((knnval-output_valid)**2)
    rss_all.append(knnval)

best_k=rss_all.index(min(rss_all))+1
print(best_k)
```

```
9
```

Comparing Differnet models

In [144]:
```
kvals = range(1, 16)
plt.plot(kvals, rss_all,'bo-')
```

Out[144]:
```
[<matplotlib.lines.Line2D at 0x7f14c0625b70>]
```



IV. CONCLUSION

- We would like to conclude this by mentioning the advantages of this model and also the easiest way to build it
- House price prediction can help sellers to determine the selling price of the property and can help the customer to purchase the right property at the right price.
- No need of chasing any agent about knowing the area demand and also infrastructure demand.
- The buyer can also get clarified whether the agent or seller is demanding the right and meaningful price or not.
- Coming to algorithms, KNN regression is one of the best ones to get the right price. It is just about making the average of k similar houses of the query house.
- Even in the real world, the price of a house is nearer to some of the similar houses. So, the same logic is applied in this algorithm which gives the output with more efficiency. Here ends the description of our project and hopefully, this is very useful.

V. REFERENCES

1. ieeexplore.ieee.org“Prediction Using Regression Techniques.”
2.An article by Victor Roman“Machine Learning Project: Predicting Boston House Prices with Regression”.
3. An article by Weikun Zhao, Cao Sun, Ji Wang,“The Research on Price Prediction of Second-hand houses
4. An article by Uday Deo,“Various Regression Techniques.”
5.(IJACSA) International Journal Of Advanced Computer Scinence And Applications. ” Modeling Using Regression Analaysis.”
6.Thesis by Venelin Valkov“Predicting House Prices with Linear Regression | Machine Learning from Scratch”