

A Report on

HOUSE PRICING PREDICTION SYSTEM

Under the Guidance (Mentor) of

Aishwarya Saxena

By – Keerthana.R

Lavanya Karanth

Ksheeraja K

Bhavana S B

Machine Learning - Python

College - NMAM Institute of Technology

Summer Internship –February- March 2021



**INDEX**

|  |  |  |
| --- | --- | --- |
| SR NO. | TOPIC | PAGE NO. |
| 1 | INTRODUCTION | 3 |
| 2 | OBJECTIVE | 4 |
| 3 | BACKGROUND | 5 |
| 4 | HARDWARE AND SOFTWARE REQUIREMENTS | 6 |
| 5 | COADING AND OUTPUT | 8 |
| 5 | FUTURE SCOPE | 15 |
| 6 | CONCLUSION | 16 |
| 7 | REFERENCES AND BIBLIOGRAPHY | 17 |

**INTRODUCTION:**

Python is a general-purpose, and high-level programming language which is best known for its efficiency and powerful functions. Python is loved by data scientists because of its ease of use, which makes it more accessible. Python provides data scientists with an extensive tools and packages to build machine learning models. One of its special features is that we can build various machine learning with less-code.

A housing market can be understood as any market for properties which are negotiated either directly from their owners to buyers, or through the services of real state brokers. People and companies are drawn to this market, which presents many profit opportunities that come from housing demands worldwide. These demands are influenced by several factors, such as demography, economy, and politics. For this reason, the analysis of such markets has been challenging data scientists and ML engineers around the world, as they must take into account a wide range of scientific fields, each one addressing different kinds of data, to come up with accurate results to customers and stakeholders The prediction of housing prices as they vary through time and place is a complex task, particularly when the available data is massive and divided into many different datatypes.

The machine will basically remember the past e-mails that needed been named similarly as spam e-mails by the human user. When another email arrives, the machine will look for it in the past set about spam e-mails. Though it matches among them, it will be trashed. Otherwise, it will make moved of the user’s inbox organizer.

Responsibilities further than Human Capabilities: an additional totally crew about errands that profit starting with machine Taking in systems are identified with the Investigation for extremely substantial and intricate information sets: galactic data, turning restorative chronicles under restorative knowledge, climate prediction, and dissection of genomic data, Web search engines, Also electronic trade. With an increasing amount accessible digitally recorded data, it gets evident that there would treasures about serious majority of the data covered clinched alongside information chronicles that would best approach excessively little and also as well perplexing to people with bode well about. Taking in with recognize serious examples over substantial Also complex information sets may be a guaranteeing space for which the blending of projects that take for the Just about boundless memory limit and expanding transforming velocity about PCs opens up new horizons.

**OBJECTIVES:**

Machine learning plays a major role from past years in image detection, spam reorganization, normal speech command, product recommendation and medical diagnosis. Present machine learning algorithm helps us in enhancing security alerts, ensuring public safety and improve medical enhancements. Machine learning system also provides better customer service and safer automobile systems. In the present paper we discuss about the prediction of future housing prices that is generated by machine learning algorithm. For the selection of prediction method, we compare and explore various prediction methods. We utilize lasso regression as our model because of its adaptable and probabilistic methodology on model selection. Our result exhibit that our approach of the issue needs to be successful, and has the ability to process predictions that would be comparative with other house cost prediction models. More over on other hand housing value indices, the advancement of a housing cost prediction that tend to the advancement of real estate policies schemes. This study utilizes machine learning algorithms as a research method that develops housing price prediction models. We create a housing cost prediction model In the view of machine learning algorithm models for example, XGBoost, lasso regression and neural system on look at their order precision execution. We in that point recommend a housing cost prediction model to support a house vender or a real estate agent for algorithm, in view of accuracy, reliably outperforms alternate models in the execution of housing cost prediction.

**BACKGROUND:**

The model is performed to identify statistically significant factors which can be able to predict the house pricing.

Here the house pricing study is done by analysing some data by giving some queries which is related to the study. And the steps involved are,

**Phase 1: Collection of data Data:**

processing techniques and processes are numerous. We collected data for Mumbai’s real estate properties from various real estate websites. The data would be having attributes such as Location, carpet area, built-up area, age of the property, zip code, etc. We must collect the quantitative data which is structured and categorized. Data collection is needed before any kind of machine learning research is carried out. Dataset validity is a must otherwise there is no point in analysing the data.

**Phase 2: Data pre-processing Data**

pre-processing is the process of cleaning our data set. There might be missing values or outliers in the dataset. These can be handled by data cleaning. If there are many missing values in a variable we will drop those values or substitute it with the average value.

**Phase 3: Training the model**

Since the data is broken down into two modules: a Training set and Test set, we must initially train the model. The training set includes the target variable. The decision tree regressor algorithm is applied to the training data set. The Decision tree builds a regression model in the form of a tree structure. Phase 4: Testing and Integrating with UI The trained model is applied to test dataset and house prices are predicted. The trained model is then integrated with the front end using Flask in python.

The language that will help us writing the program is Python.

Python contains library such as Numpy, Panda, Scikit-Lean, Scipy, from pandas.tools.plotting import scatter\_matrix, import matplotlib.pyplot as plt.

**HARDWARE AND SOFTWARE REQUIREMENTS:**

**Hardware Requirements:**

|  |  |
| --- | --- |
| **Hardware Tools** | **Minimum Requirements** |
| Processors | i5 or above |
| Hard Disk | 10GB |
| RAM | 8GB |
| Monitor | 17’’ Colored |
| Mouse | Optical |
| Keyboard | 122 Keys |

**Software Requirements:**

|  |  |
| --- | --- |
| **Software Tools** | **Minimum Requirements** |
| Platform | Windows, Linux or MacOS |
| Operating System | Windows, Linux or MacOS |
| Technology | Machine Learning-Python |
| Scripting Language | Python |
| IDE | Pycharm (and Sublime) |

**CODING AND OUTPUT:**

**import** numpy **as** np

**import** pandas **as** pd

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

**import** mpl\_toolkits

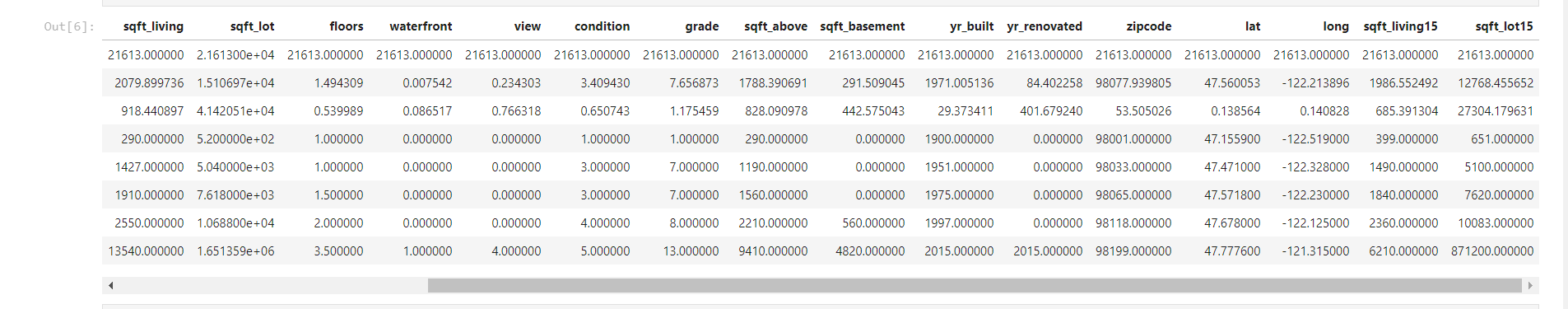
**%matplotlib** inline

data **=** pd**.**read\_csv("kc\_house\_data.csv")

data**.**head()

data**.**describe()

**Output:**

****

**Table. 1**

data['bedrooms'].value\_counts().plot(kind='bar')

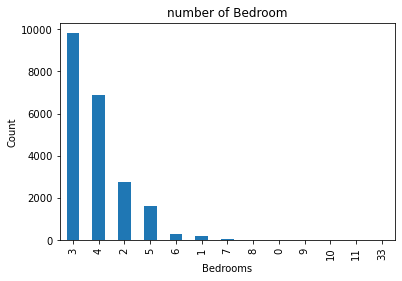
plt.title('number of Bedroom')

plt.xlabel('Bedrooms')

plt.ylabel('Count')

sns.despine

**Output:**

****

**Fig.2. Number of bedrooms**

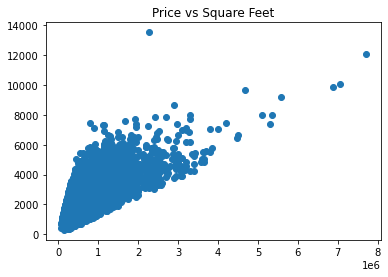
plt**.**scatter(data**.**price,data**.**sqft\_living)

plt**.**title("Price vs Square Feet")

plt**.**show()

sns**.**despine()

**Output:**



**Fig.3. Price vs square feet**

plt.scatter(data.price,data.long)

plt.title("Price vs Location of the area")

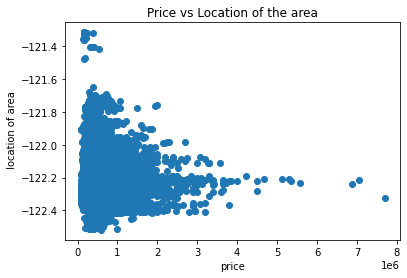
plt.xlabel('price')

plt.ylabel(' location of area')

plt.show()

sns.despine()

**Output:**



**Fig.4. Price vs Location of the area**

plt**.**scatter(data**.**price,data**.**lat)

plt**.**xlabel("Price")

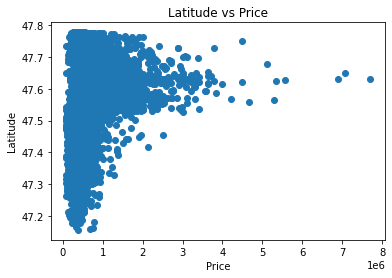
plt**.**ylabel('Latitude')

plt**.**title("Latitude vs Price")

plt**.**show()

sns**.**despine()

**Output:**



**Fig.5. Latitude vs Price**

plt**.**scatter(data**.**bedrooms,data**.**price)

plt**.**title("Bedroom and Price ")

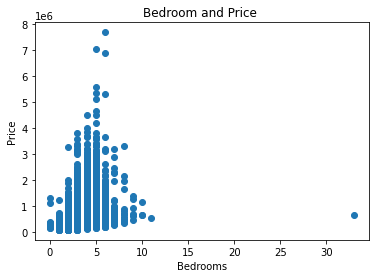
plt**.**xlabel("Bedrooms")

plt**.**ylabel("Price")

plt**.**show()

sns**.**despine()

**Output:**



**Fig.6. Bedroom ad price**

plt**.**scatter((data['sqft\_living']**+**data['sqft\_basement']),data['price'])

plt**.**title('Price vs sqft area')

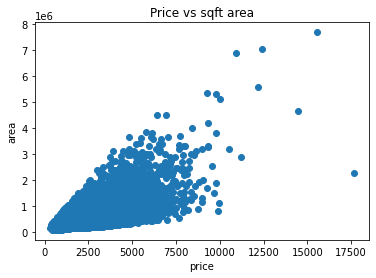
plt**.**xlabel('price')

plt**.**ylabel('area')

plt**.**show()

sns**.**despine()

**Output:**



**Fig.7. Price vs sqft area**

plt.scatter(data.waterfront,data.price)

plt.title("Waterfront vs Price ( 0= no waterfront)")

plt.xlabel('Waterfront')

plt.ylabel('Price')

plt.show()

sns.despine()

**Output:**



**Fig.8. Waterfront vs price**

plt**.**scatter(data**.**condition,data**.**price)

plt**.**title("condition vs price")

plt**.**xlabel('condiotion')

plt**.**ylabel('Price')

plt**.**show()

sns**.**despine()

**Output:**

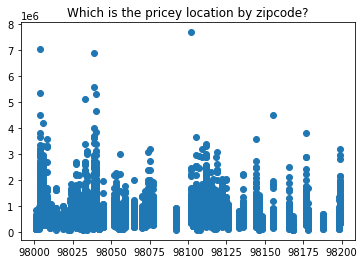


**Fig.9. Condition vs price**

plt**.**scatter(data**.**zipcode,data**.**price)

plt**.**title("Which is the pricey location by zipcode?")

**Output**



**Fig.10. Pricey location by zipcoad**

**FUTURE SCOPE:**

In the future, we are going to present a comparative study of the systems’ predicted price and the price from real estate websites such as Housing.com for the same user input. Also, to simplify it for the user, we are going to recommend real estate properties to the user based on the predicted price. The current dataset only includes cities of Mumbai, expanding it to other cities and states of India is the future goal. To make the system even more informative and user-friendly, we will be including Gmap. This will show the neighborhood amenities such as hospitals, schools surrounding a region of 1 km from the given location. This can also be included in making predictions since the presence of such factors increases the valuation of real estate property.

**CONCLUSION:**

In today’s real estate world, it has become tough to store such huge data and extract them for one’s own requirement. Also, the extracted data should be useful. The system makes optimal use of the Linear Regression Algorithm. The system makes use of such data in the most efficient way. The linear regression algorithm helps to fulfill customers by increasing the accuracy of estate choice and reducing the risk of investing in an estate. A lot’s of features that could be added to make the system more widely acceptable. One of the major future scopes is adding estate database of more cities which will provide the user to explore more estates and reach an accurate decision. More factors like recession that affect the house prices shall be added. In-depth details of every property will be added to provide ample details of a desired estate. This will help the system to run on a larger level.

**REFERENCES AND BIBLIOGRAPHY:**

1. Atharva chogle , priyanka khaire , Akshata gaud , Jinal Jain. House Price Forecasting using Data Mining Techniques

2. Nihar Bhagat, Ankit Mohokar, Shreyash Mane. House Price Forecasting using Data Mining.

3. G. Naga Satish, Ch. V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu. House Price Prediction Using Machine Learning

4. Datasets ([www.kaggle.com](http://www.kaggle.com))

5. Sklearn libraries (<http://scikit-learn.org/>)

6. Algorithms references

* <https://towardsdatascience.com/>
* <https://medium.com/codex/house-price-prediction-with-machine-learning-in-python-cf9df744f7ff>
* <https://medium.com/analytics-vidhya/predicting-house-prices-using-classical-machine-learning-and-deep-learning-techniques-ad4e55945e2d>