House Price Prediction

Harpreet Kaur
Department (ECE)
Chandigarh University, Mohali, India

Kriti Gupta
Department (ECE)
Chandigarh University, Mohali, India

Abstract-*House Price Index (HPI) is regularly used to gauge the progressions in lodging cost. Since lodging cost is unequivocally related to different factors, for example, area, region, populace, it requires other data separated from HPI to foresee individual lodging cost. There has been an impressively huge number of papers taking on customary AI ways to deal with anticipate lodging costs precisely, however they seldom worry about the presentation of individual models and disregard the less famous yet complex models. Thus, to investigate different effects of highlights on forecast techniques, this paper will apply both customary and high-level AI ways to deal with examine the distinction among a few high-level models. This paper will likewise thoroughly approve various methods in model execution on relapse and give a hopeful outcome to lodging cost forecast.*
Keywords: *House Price Index, Artificial Intelligence, Machine Learning*

I.   Introduction

In India, lodging goes from previous maharajas' castles to present day high rises in significant urban communities to little hovels in far off regions. As income have developed, India's lodging industry has seen critical extension. Buying a house is a tedious and bewildering process. According to the report, the typical lodging cost in a metro is Rs15 lakh. It additionally exhibits that purchasing a home in a specialty city is more costly than purchasing a home in a blast city.

The middle property cost in specialty urban communities is Rs10 lakh, contrasted with Rs8 lakh in boomtowns. Land will be land, structures, and basic resources like harvests, climatic condition, and water, sewer lines: an interest invested in this (likewise) a thing of genuine property, structures or houses in general. Personal property, like vehicles, boats, gems, furniture, hardware, and farming moving stock, is unmistakable from land since it isn't for all time attached to the ground.

Private land can be a solitary family or multifamily structure that is accessible for home or non-business use[1]. How homes are associated with different homes and land can be utilized to group them. For similar actual sort, many types of lodging tenure may be utilized. Associated condos, for instance, may be held by one business and rented out, or they could be possessed freely with an arrangement framing the connection among units and shared spaces and concerns.

There are different classes and the sorts of houses in it which are:
- Appended/multi-unit abodes
- Condo
- multi-family house
- Terraced house
- Helpful
- Semi-isolates homes
- Duplex
- Isolates homes
- Isolates house or single-family withdrawn house
- Versatile homes
- Manufactured homes or private convoys
- Houseboats
- Tents

Land is often obtained as an interest in regions where land and development costs are rising, whether the proprietor intends to utilize the property. Frequently, venture homes are leased, however "flipping" involves

quick exchanging a property, once in a while to exploit exchange or quickly developing worth, and some of the time after critical redesigns have been finished. Extravagance land is once in a while used for of putting away resources, especially by rich outsiders, without any goal of leasing it out.

The asking cost and essential portrayal are regularly introduced autonomously from the nonexclusive and normalized land credits. These elements might be effectively analyzed across the entire range of potential homes since they are given separately and in an orderly way. Property sellers can include a summary of all the main qualities of the house in the description because each house has its own unique traits, such as a specific view or style of sink. Potential purchasers can evaluate all real estate qualities, but owing to the vast diversity, providing an automatic comparison on all variables is almost difficult. This is also true in the opposite direction: home sellers must estimate the worth of their property based on its qualities in comparison to similar properties currently on the market.

As the fact suggests there are various qualities and features which are humanly is not possible for the real estate agent to show the buyers or to even remember in his/her memory. Because of which, he only tends to remember only the highlights of the properties such as:

- The area of the house
- Information about the surroundings for example schools, malls, markets etc
- The house's flooring material
- The kitchen dimensions
- How many washrooms does the house consists
- Does the house have a terrace
- The electricity supplies the water supply whether it is for specific timings or for 24 hours.

The highlights tend to create a fake image in the customers mind and sometimes are used to hide the small things which are important but not discussed. Due to which the customers result in regretting buying the property after a few years. When they encounter those lacks and those small things which were not discussed and were important then it becomes a pain. We as a team understood the importance of every small aspect which makes a perfect property to live in and to work in. Then we came up with a project called 'House Price Prediction using Machine Learning'. This can help not just the customers, but also the sellers of the real estate.

To get the account of such an enormous amount of data which consists of almost all the factors ranging from smallest to the biggest. We used machine learning as a tool. because it is capable of processing vast amounts of data in a properly sorted manner by its various techniques and algorithms[2]. Our topic works to remove that stress of the buyers and the sellers and also in creating transparency while buying or selling the property.

II. Limitations Of Methodologies

There are some astonishing measures of study done over house cost forecast office yet limited work has showed up to any authentic instrumentation. There is practically no confirmation of a working house cost marker set up by an association. For now, not a lot of cutting-edge game plans exist for such a huge market and an enormous piece of the techniques used by people and associations are according to the accompanying:

Purchasers/Customers:

At the point when individuals first consider purchasing a house/Real space they will frequently go on the web and endeavor to focus on designs and some other associated stuffs. People do this so they can look for a house which contains all amenities that they require. While doing this, people make a note of the expenses related to with the houses. Anyhow, a typical individual does not have any accurate idea and sufficient data about what the actual cost of the houses should be. This can often leads to fraud and trickery as the accepts everything the saw on different websites from the internet and believed it to be certified.

Another major thing that popped up at the time of searching for a property is to contact different Estate consultants. The complication with these estate consultants is that they wanted to get paid a proportion for searching a house and then they set the price tag for you. All around, this retail cost is capriciously acknowledged by people since they doesn't have any other option. There might be cases that these consultants and vendors could have a secret overseeing and the client might be sold an exaggerated house without the customer's understanding.

Venders/Companies:

Whenever a person decides selling their properties they differ their property and numerous of different properties which are posted from one side of the planet to the other. Deciding the cost by contrasting it and various homes is profoundly tedious and has a likely gamble of inaccurate estimating.

Large Real home organizations have different items they need to sell and they need to downgrade individuals in order to get all these items sold. It once again grounds the conjecture of a retail cost on a human thusly there is space for human misstep. Likewise, these downgraded individuals ought to get paid. In spite of these facts, a PC can get this work done for you by working through the weighty figures that can save a lot of time and money. Also it will give the precision that a human can't accomplish[3].

III. Literature Review

AI (ML) is an investigation of PC calculations that can naturally further develop data and information utilization. It is viewed as a component of man-made reasoning. AI calculations are utilized in a wide range of projects, like clinical, self-driving vehicles (Tesla and so forth), and PC vision, where it is troublesome or difficult to foster normal calculations to play out the expected errands. Regulated learning has stacks of data and when you show the data to a machine, the machine stores the data inside itself. Land has become in excess of a need in this 21st hundred years, it addresses something considerably more these days. Not just for individuals investigating purchasing Real Estate yet additionally the organizations that sell these Estates. As per Real Estate Property isn't just the essential need of a man however today it likewise addresses the wealth and glory of an individual. Interest in land by and large is by all accounts productive on the grounds that their property estimations don't decline quickly. Changes in the land cost can influence different family financial backers, investors, policymakers, and many. Interest in the land area is by all accounts an appealing decision for speculations. Accordingly, foreseeing the land esteem is a significant monetary record. Proposes that each and every association in the present land business is working productively to accomplish a strategic advantage over elective contenders. There is a need to improve on the cycle for a typical individual while giving the best outcomes. Proposed to utilize AI and man-made consciousness strategies to foster a calculation that can foresee lodging costs in view of specific info highlights. The business utilization of this calculation is that characterized sites can straightforwardly utilize this calculation to anticipate costs of new properties that will be recorded by taking a few information factors and foreseeing the right and supported cost i.e., try not to take cost inputs from clients and accordingly not letting any blunder sneaking in the framework. We have done the implementations on Google Colab, which is an open-source web application that allows everybody to write and run codes, there we can get all the python libraries and that's why it's best suited for machine learning and analyzing the data. It contains apparatuses for information cleaning, information change, reenactment of numeric qualities, displaying utilizing insights, representation of information, and AI devices. We planned a framework that will assist individuals with knowing near the exact cost of land[4]. Clients can give their essentials as specified by them by which they will get the expenses of the ideal house. Purchasers can similarly get the model arrangement of the house to get the recommendations for the houses.

IV. Methodology

We have utilized information pre-handling for the forecast of house costs. Information preprocessing is a course of setting up the crude information and making it reasonable for an AI model. It's a fundamental stage while making an AI model. This present reality information by and large contains commotions, missing qualities, blunders and perhaps in unusable arrangement which can't be straightforwardly utilized for making models so that is the reason every one of these irksome datasets are taken out utilizing information pre-handling which additionally builds the precision and effectiveness of our AI model[5]. Information pre-handling is isolated into four classes called a) Data cleaning b) Data altering c) Data deleting and d) Data fighting.

Data cleaning is the most common way of recognizing and adjusting bad or mistaken records from a dataset or distinguishing fragmented or insignificant pieces of the information and afterward supplanting and altering the information.

Data altering is an interaction where anomalies are picked from information and annihilated. Exceptions are principally kept in information primarily because of exploratory mistakes created by machines due to breaking down or because of a few different boundaries.

Data deleting is a strategy where information is diminished to its least difficult structure to let loose limit on a capacity gadget. Invalid information are diminished utilizing a portion of the standardization strategies. Z score is one of the cycles utilized for standardization.

Data fighting is the most common way of changing and planning information from crude information structure into one more organization with the aim of making it more suitable and important for the reason[6]. The most common way of fighting likewise incorporates munging, information representation, information accumulation, and preparing a measurable model. Information representation produces diagrams for the factual model and information conglomeration is the assembling of data from datasets for information handling.

V. Implementation

Datasets

We have utilized the dataset "MagicBricks" and applied the Linear Regression models for the expectation. We have taken the datasets from Kaggle.com site which assists the clients with finding the datasets, investigate and assemble models in an electronic information science climate. Our code is executed on Google Colab using Python language. We have involved different python libraries in this pathway, for example, Pandas, Scikit Learn, Seaborn, Numpy and Matplotlib. Our model depends on Supervised Learning, which is a subcategory of Machine Learning. Regulated Learning utilizes named datasets to prepare calculations to characterize information or anticipate the results precisely. The model which we have utilized in our task is Linear Regression and foreseeing the worth of a variable in light of the worth of another variable is utilized[7]. The variable which we need to foresee is called subordinate variable and the variable we are utilizing to anticipate the worth of ward variable is called free factor. This model makes the assessment technique basic and direct conditions are straightforward. The model of direct relapse incorporates information investigation, information perception, information determination, information change, preparing of datasets, testing of datasets and afterward the last expectation is finished.

Data Exploration

It is the initial step of information examination used to investigate and envision information to reveal bits of knowledge from the beginning or recognize regions or examples to dive into more. Utilizing intelligent dashboards and point-and-snap information investigation, clients can more readily figure out the master plan and get to experiences quicker.
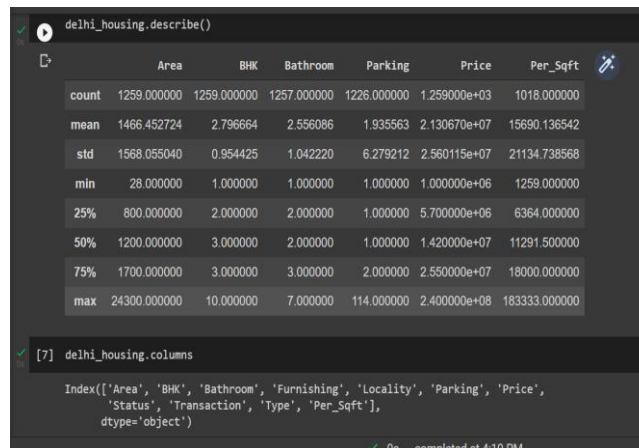

Figure 1. Data Exploration

Data Visualization

It is the portrayal of information through utilization of normal designs, like graphs, plots, infographics, and even liveliness. These visual showcases of data impart complex information connections and information driven experiences in a manner that is straightforward.
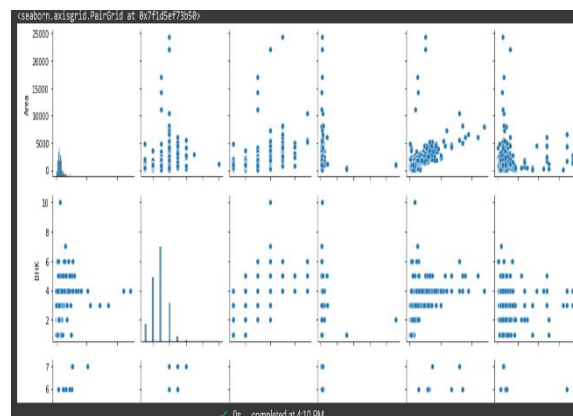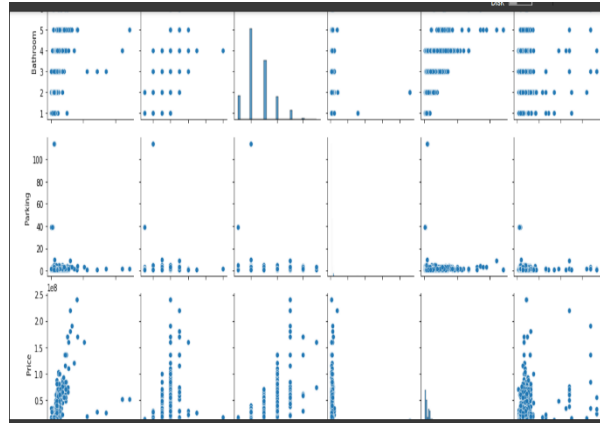

Figure 2. scatter plot of dataset

Figure 3. scatter plot of dataset

Data Selection
It is characterized as the most common way of deciding the fitting information type and source, as well as reasonable instruments to gather information. Information determination goes before the real act of information assortment.

	Price	Area_Sqm	Bedrooms	Latitude	Longitude	PricePerSqM
0	13300000	74.32	2	19.210200	72.864891	178885.00
1	9000000	55.74	1	19.167700	72.949300	168566.16
2	9000000	43.66	1	19.130609	72.873816	206032.48
3	9000000	78.97	2	19.126005	72.825052	113926.88
4	100000000	464.51	5	19.075014	72.907571	215200.00

Figure 4. Selecting appropriate data

Data Transformation
It is the most common way of changing over information from one arrangement, for example, a data set record, XML archive or Excel bookkeeping sheet, into another. Changes commonly include changing over a crude information source into a purged, approved and prepared to-utilize design[8].

Training the datasets
The preparation information is an underlying arrangement of information used to assist a program with understanding how to apply innovations like brain organizations to learn and deliver complex outcomes. It very well might be supplemented by resulting sets of information called approval and testing sets.


```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Activation
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.layers import Dropout
from tensorflow.keras.callbacks import EarlyStopping
early_stop = EarlyStopping(monitor='val_loss', mode='min', verbose=1, patience=25)

model = Sequential()
model.add(Dense(371,activation='relu'))
model.add(Dropout(0.5))

model.add(Dense(185,activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(93,activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(46,activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(1))

model.compile(optimizer='adam',loss='mse')
```

Figure 5. Training of dataset

Testing the datasets

A straightforward assessment strategy is a train test dataset where the dataset is partitioned into a train and a test dataset, then the learning model is prepared utilizing the train information and execution is estimated utilizing the test information. In a more complex methodology, the whole dataset is utilized to prepare and test a given model.
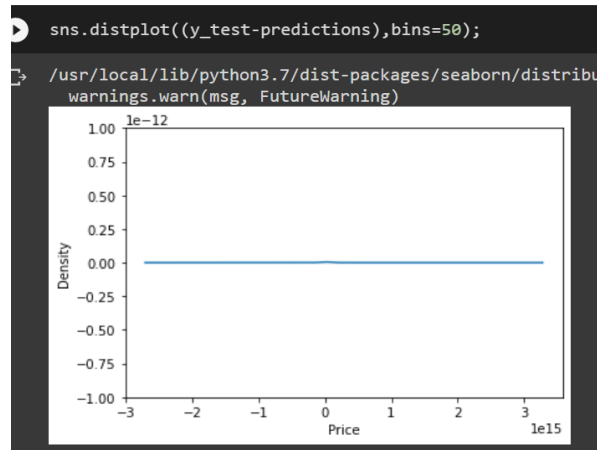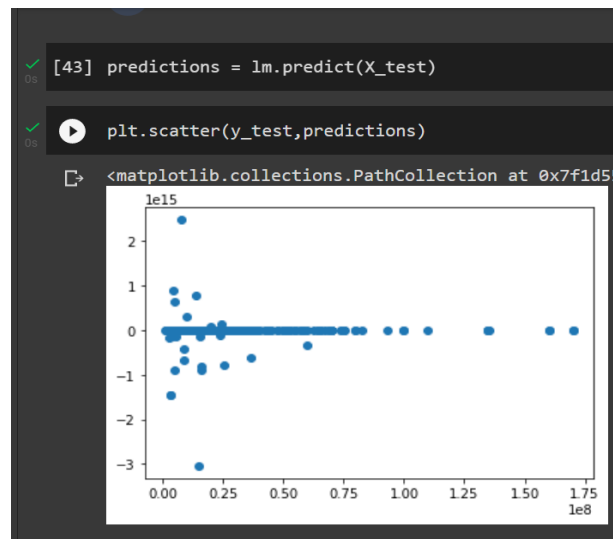

Figure 6. graph of tested data

Prediction

Figure 7. final prediction model testing

VI. Results

The proposed framework is executed on google colab utilizing python language and the datasets are taken from Kaggle contest on house cost forecast. In this venture, we have fabricated an AI model to foresee the house costs of an Indian city and expectation models have effectively accomplished an incredible scope of exactness from 70% to 80%. The models which we took being used were[9]:

• Linear Regression : It is an AI model which tracks down the relationship between's one free item and one ward object by giving the best fit straight line. This model gave the most elevated precision of 85.64% during the run time.

• Decision tree : It is a non-parametric realizing which requires oversight and its utilization in the program is to make a model which can anticipate the worth of the designated variable by making choice standards from the information highlights. This had minimal precision in the program.

• Vector regression: It is likewise a piece of realizing which requires management and it predicts the discrete qualities by means of utilizing its calculations and in the code it gives the precision level at 73%.

This task will be exceptionally useful for the housing market. Our model can be utilized by both house venders and house purchasers. We utilize different relapse procedures in this pathway and our outcomes are not

the sole assurance of one method rather it is the weighted mean of different strategies to give the most reliable outcomes[10]. The outcome demonstrated that this approach yields least mistake and most extreme precision that singular calculations applied. As you can see below:
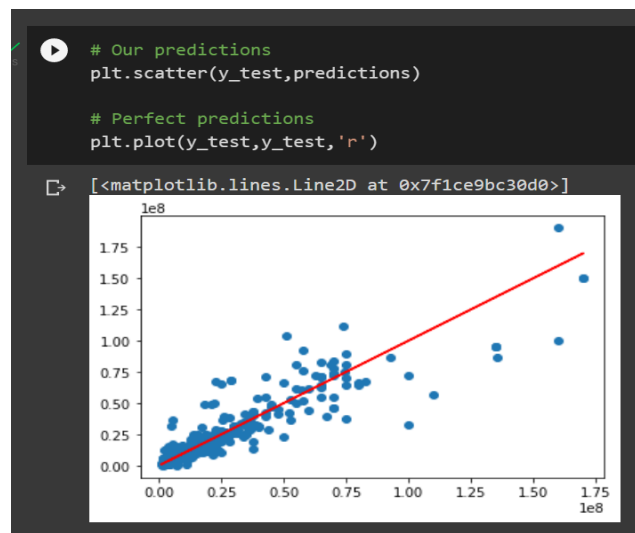

Figure 8. graph of house pricing

Figure 9. prediction of single house pricing

VII.   Conclusion and Future scope

Every individual wishes for purchasing their own home. By using this proposed model, we consider individuals to purchase houses and land at their authentic expenses and need to ensure that they do not get tricked by crude specialists who simply are after their cash. Moreover, our model will also help the big companies to set the best prices for housing and rescue them from a lot of issues and a huge amount of time and cash will get saved. Right land costs are the pith of the market and we need to guarantee that by utilizing this model[11].

The framework is well-suited sufficient in preparing itself and in foreseeing the costs from the crude information gave to it. In the wake of going through a few examination papers and various websites, a bunch of calculations were chosen which were appropriate in implementing the dataset of the model. After performing a lot of training and testing, it was determined that the Linear Regression Algorithm appeared to show the finest outcomes among the other calculations. The structure was powerful enough at predicting the price of houses with different highlights and had the option to deal with huge amounts of information. The framework is very easy to understand and efficient[12].

The beneficial component that can be added to our proposed framework is to benefit clients of an undeniable UI so that there can be more functionalities for the clients that can be utilized with machine learning models for different areas.

References

[1].   Truong, Q., Nguyen, M., Dang, H., & Mei, B. (2020). Housing price prediction via improved machine learning techniques. *Procedia Computer Science*, *174*, 433-442.

[2]. Thamarai, M., & Malarvizhi, S. P. (2020). House Price Prediction Modeling Using Machine Learning. *International Journal of Information Engineering & Electronic Business*, *12*(2).

[3]. Madhuri, C. R., Anuradha, G., & Pujitha, M. V. (2019, March). House price prediction using regression techniques: a comparative study. In *2019 International Conference on Smart Structures and Systems (ICSSS)* (pp. 1-5). IEEE.

[4]. Jiang, Z., & Shen, G. (2019, November). Prediction of house price based on the back propagation neural network in the keras deep learning framework. In *2019 6th International Conference on Systems and Informatics (ICSAI)* (pp. 1408-1412). IEEE.

[5]. Wu, J. Y. (2017). Housing price prediction using support vector regression.

[6]. Jain, M., Rajput, H., Garg, N., & Chawla, P. (2020, July). Prediction of house pricing using machine learning with python. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 570-574). IEEE.

[7]. Kuvalekar, A., Manchewar, S., Mahadik, S., & Jawale, S. (2020, April). House Price Forecasting Using Machine Learning. In *Proceedings of the 3rd International Conference on Advances in Science & Technology (ICAST)*.

[8]. Manasa, J., Gupta, R., & Narahari, N. S. (2020, March). Machine learning based predicting house prices using regression techniques. In *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 624-630). IEEE

[9]. Wang, P. Y., Chen, C. T., Su, J. W., Wang, T. Y., & Huang, S. H. (2021). Deep learning model for house price prediction using heterogeneous data analysis along with joint self-attention mechanism. *IEEE Access*, *9*, 55244-55259.

[10]. Wang, P.Y., Chen, C.T., Su, J.W., Wang, T.Y. and Huang, S.H., 2021. Deep learning model for house price prediction using heterogeneous data analysis along with joint self-attention mechanism. *IEEE Access*, *9*, pp.55244-55259.

[11]. Varma, Ayush, et al. "House price prediction using machine learning and neural networks." *2018 second international conference on inventive communication and computational technologies (ICICCT)*. IEEE, 2018.

[12]. Madhuri, CH Raga, G. Anuradha, and M. Vani Pujitha. "House price prediction using regression techniques: a comparative study." *2019 International Conference on Smart Structures and Systems (ICSSS)*. IEEE, 2019.