PREDICTING HOUSE PRICE USING MACHINE LEARNING

Abstract:

Real Estate industry is dynamic in terms of the prices being fluctuated regularly. It's one of the main area to apply the machine learning concepts to predict the prices of real estate depending upon the current situations and make out maximum accuracy for the same. The research paper mainly focus on to predicting the real valued prices for the places and the houses by applying the appropriate ML algorithms. The proposed article considers some essential aspects and parameters for calculating the prices of real estate property Also some more geographical and statistical tech-inquest will be needed to predict the price of a house. The paper consist how the house pricing model works after using some machine learning techniques and algorithms. The use of the dataset in the proposed system from the reputed website helps to get the detailed analysis of the data points. Algorithms like Linear regression and sklearn are used to effectively increase the accrue-cy. During model structure nearly all data similarities and cleaning, outlier removal and feature engineering, dimensionality reduction, grid search cv for hyper parameter tuning, k fold cross-validation, etc. are covered.

House prices are a significant impression of the economy, and its value ranges are of great concerns for the clients and property dealers. Housing price escalate every year that eventually reinforced the need of strategy or technique that could predict house prices in future. There are certain factors that influence house prices including physical conditions, locations, number of bedrooms and others. Traditionally predictions are made on the basis of these factors. However such prediction methods require an appropriate knowledge and experience regarding this domain. Machine Learning techniques have been a significant source of advanced opportunities to analyze, predict and visualize housing prices. In this paper, Gradient Boosting Model XGBoost is utilized to predict housing prices. Publicly available dataset containing 38,961 records of Karachi city is attained from an Open Real Estate Portal of Pakistan.

Introduction:

Development of civilization is the foundation of the increase in demand for houses day by day. Accurate prediction of house prices has been always a fascination for buyers, sellers, and bankers also. Many researchers have already worked to unravel the mysteries of the prediction of house prices. Many theories have been given birth as a consequence of the research work contributed by various researchers all over the world. Some of these theories believe that the geographical location and culture of a particular area determine how the home prices will increase or decrease whereas other schools of thought emphasize the socio-economic conditions that largely play behind these house price rises.

We all know that a house price is a number from some defined assortment, so obviously prediction of prices of houses is a regression task. To forecast house prices one person usually tries to locate similar properties in his or her neighborhood and based on collected data that person will try to predict the house price.

All these indicate that house price prediction is an emerging research area of regression that requires the knowledge of machine learning. This has motivated me to work in this domain.

Real-estate appraisal is an integral part of the property buying process. Traditionally, the appraisal is performed by professional appraisers specially trained for real estate valuation. For the buyers of real estate properties, an automated price estimation system can be useful to estimate the prices of properties currently on the market. Such a system can be particularly helpful for novice buyers who are buying a property for the first time, with little to no experience.

Data collection:

Data processing techniques and processes are numerous. We collected data for USA/Mumbai 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, price, no of bedrooms 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 analyzing the data.

Data processing:

Data preprocessing 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.

Attributes:

housing sector. Residents' quality of life as well as national economy depends on the potential house price increase.

Ultimately, this issue will affect investors who are making their house as an investment. An increase in house demand occurs each year, indirectly causing house price increases every year. The problem arises when there are numerous variables such as location and property demand that may influence the house price, thus most stakeholders including buyers and developers, house builders and the real estate industry would like to know the exact attributes or the accurate factors influencing the house price to help investors make decisions and help house builders set the house price. House price prediction can be done by using a multiple prediction models (Machine Learning Model) such as support vector regression, artificial neural network, and more. There are many benefits that home buyers, property investors, and house builders can reap from the house-price model. This model will provide a lot of information and knowledge to home buyers, property investors and house builders, such as the valuation of house prices in the present market, which will help them determine house prices. Meanwhile, this model can help potential buyers decide the characteristics of a house they want according to their budget [5]. Previous studies focused on analyzing the attributes that affect house price and predicting house price based on the model of machine learning separately. However, this article combines such a both predicting house price and attributes together. In this article, literature review focuses on predicting house price based on the model of machine learning as well as analyzing attributes primarily used in previous study that affect house price. This paper was arranged as follows:

the first section summarizing overall of this study. Second section described the common attributes used in prediction of house price around the world. It was followed by a brief discussion of machine learning model used in previous study to predict house price. For the next section, the comprehensive effects of the current house price prediction model are addressed. Ultimately, section 5 and section 6 respectively provide the description and conclusion of this comprehensive literature analysis.

2. Attributes:

House price prediction can be divided into two categories, first by focusing on house characteristics, and secondly

by focusing on the model used in house price prediction. Many researchers have produced a house price prediction
model, including [1, 3, 6–8].
A research undertaken by [9] analyses the existing housing price in Jakarta, Indonesia using the conceptual model
and questionnaires. Based on the results, the attributes or factors affecting the house price differ for each house
construction in Jakarta, therefore accepting the validity of this analysis as the main purpose of this research is to classify
the factor or attributes affecting the house price. Various considerations influence the price of a house. According to
[10], the factors influencing house prices can be classified into three categories: location, structural and neighborhood condition.

A. Location:

location is considered to be the most significant feature of house price determination [6, 9–11]. [12] in his study also observed the significant of location attributes in deciding house price. The location of the property was classified in a fixed vocational attribute. All of these studies point to the close association between locational attributes such a distance from the closest shopping center, or position offering views of hills or shore, and house price variation.

B. Structural:

Another significant feature influencing the house price is structural structure or some research has listed it as physical attributes [10, 13]. Structural characteristic is a feature that people may identify, whether number of bedrooms and bathrooms, or floor space, or garage and patio. These structural attributes, often offered by house builders or developers to attract potential buyers, therefore meet the potential buyers' wishes. In his earlier study, structural attributes would be the key consideration for house hunters in determining what to purchase as such attributes represent their market value. In their earlier study, stated that all these attributes have a positive relationship to rising house price.

C. Neighborhood:

Neighborhood qualities can be included in deciding house price. According to [13], efficiency of public education, community social status and proximity to shopping malls typically improve the worth of a property. There is a substantial rise in house prices from the fifth-class suburban community to affluent neighborhood as predicted [16]. Nonetheless, [13]

study found that these qualities tend to be cultural based, as they are not similarly relevant in all cultures.

Related Work:

A total of 14 articles were reviewed and evaluated to capture all attributes that influences the price of house. [3] in his article stated that square footage of unit of a house is the most importance variable in predicting price of a house, followed by number of bathrooms and number of bedrooms. Apart from that, the study suggests that the worth of the house increases by 2.6% if the floor space of the house is raised by 100 square feet. They also conclude that when the building's operating year is 1 year lower, the value increase by 0.3 percent. In addition, the price of a house would increase by 10.4 or 13.7 percent, with one more bedroom or one more bathroom.

Program:

```
 from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
import pandas as pdd

# Loading the dataset
data_h = pdd.read_csv('kc_house_data.csv')

# Selecting the features and target variable
Features1 = ['bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot', 'floors',
'zipcode']
target = 'price'
X1 = data_h[features1]
y1 = data_h[target]

# We will perform the data splitting into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X1, y1, test_size=0.2,
random_state=42)

# instance of the Linear Regression model creation
model = LinearRegression()

# Training the model
model.fit(X_train, y_train)

# Making predictions on the test set
y_pred = model.predict(X_test)

# Evaluating the model
score = model.score(X_test, y_test)
print("Model R^2 Score:", score)
# Predicting the price of a new house
new_house = pdd.DataFrame({'bedrooms': [2], 'bathrooms': [2.5],
'sqft_living': [600], 'sqft_lot': [600], 'floors': [2], 'zipcode': [98008]})
predicted_price = model.predict(new_house)
print("Predicted Price:", predicted_price[0])
```

Output:

```
C:\Users\Tutorialspoint>python image.py
Model R^2 Score: 0.5152176902631012
Predicted Price: 121215.61449578404
```

Conclusion:

In conclusion, using machine learning in Python is a powerful tool for predicting house prices. By gathering and cleaning data, visualizing patterns, and training and evaluating our models, we can make informed decisions in the dynamic world of real estate.

By leveraging advanced algorithms and data analysis, we can make accurate predictions and inform decision-making processes. This approach empowers buyers, sellers, and investors to make informed choices in a dynamic and competitive market, ultimately maximizing their opportunities and outcomes.