HOUSE PRICE FORECASTING USING MACHINE LEARNING

AmanGoyal
Department of Computer ScienceandEngineering
Galgotias University
Greater Noida,India

UtkarshKumar
Department of Computer Science and Engineering
Galgotias University
Greater Noida,India

Abstract — The real estate market is one of the world's most price-sensitive, and it is always changing. It's one of the most critical fields where computer technology can be used. exactitude in learning how to improve and predict costs in high-risk situations The aim of the paper is to make a future prediction. This system is helps in finding the starting price of a property based on geographical factors Through analysing past data patterns and price ranges in the industry, as well as future developments, The potential costs will be calculated. The scope of this investigation is as follows: You can forecast house prices in Mumbai using a decision tree.

Keywords—Decision tree regressor,machine learning.

I. INTRODUCTION

Today, every real estate company is working hard to gain a competitive advantage over its rivals. There is a requirement to make the process more user-friendly while still delivering the best results . The regression machine learning algorithm is used in this paper to suggest a strategy for predicting house prices. If you want to sell a house, you'll need to figure out how much you can ask for it.

A machine measurement would also provide you with an accurate gauge! This regression model is intended to predict not only the price of a house that is ready to sell, but also the price of a house that is still being built.

Regression is a machine learning methodology that analyses the associations among your target parameter and a number of independent factors to create projections based on current measurable data. The cost of a house is determined by factors such as the number of rooms, living space, and location, according to this theory. We can compute house valuations in a particular land area if we apply counterfeit figuring out how to these requirements.

In this proposed model, the target function is the price of the real estate property, with the following independent features: amount of rooms, bathrooms, carpeting, developed area, grounds, land age and postcode.

. Two new features - pollution levels and violence rate - were introduced to the previously listed features, which are usually used for estimating house prices.

- . Higher values of these features suggest a decline in house prices, making them useful in predicting home prices.

The Python programming language is used in the software. The predictive model is built using a Decision tree regressor from the "Scikit-learn" machine learning library. Grid Search CV will help you choose the best max-depth value for your decision tree.Flask is used to deploy the model (a pythonframework).

II. RELATED WORK

The infrastructure facilities that surround a property decide its worth. Different innovations have recently come along with a few authors' scopes for finding the best properties for customers.

Raghunandhan went through the basics of data mining, including how it works and the algorithms that help prediction. The most important step is to figure out which machine learning algorithm is best for predicting house prices. Manjula is a source of information.

Several main features to consider when predicting property prices with high accuracy usingSince environmental factors decide what kind of price we can expect for different types of houses, we'll use a regression model. Using Google maps and real-timeneighbourhood info, A. Varma devised a method for obtaining precise real-world valuations.

Studies have identified links between a city's outer appeal and non-visual factors such as crime rates, home values, and population size, among others. For example, "City Forensics: Using Visual Elements to Predict Non-Visual City Attributes" uses visual attributes to predict the selling price of a house.

.

Hujia Yu and Jiafu Wu (2014)
[5] used classification and regression algorithms. The living area square feet, roof material, and neighbourhood have the most statistical importance in deciding a home's selling price, according to the report. The PCA technique can also help with prediction analysis.Li Li and Kai-Hsuan Chu studied various architectures.
The backpropagation neural network (BPN) and
Radial base functional neural networks (RBF neural networks) are a type of RBF neural network. The use of RBF and BPN models to classify the Cathy's distance between the house price index and the Cathy' and a sinister price index, as wellas a complex correlation macroeconomic analysis detection method. Linear regression algorithms for forecasting were investigated by NiharBhagat, AnkitMohokar, and Shreyash Mane (2016) [7].
The aim of this paper is to estimate one of

most cost-effective property price for consumers based on their expectations and expenditures. Past industry dynamics and price ranges will be used to forecast future house pricing.

II. SYSTEM DESIGNAND ARCHITECTURE

Phase 1: Collection of data

There are a plethora of data processing methods and processes from which to select. We gathered data on Mumbai real estate properties from a number of different websites. The data will include attributes such as location, carpet area, built-up area, property age, zip code, and others. Quantitative data that is organised and categorised is needed. Data must be collected before any machine learning research can begin. The dataset must be valid; otherwise, it will be useless. It's pointless to analyse the results.

Phase II: Preprocessing of data

The dataset can contain null values or anomalies.The method to analyse our dat set is known as data preprocessing.

Phase III: Training the model

Since the data is split into two modules: a Training set and a Test set, we must first train the model. The instruction set includes the goal element. The decision tree regressor algorithm is applied to the training data collection. To construct a regression model, the decision tree employs a tree structure.

Phase 4: Testing and UI Integration

On a research dataset, the learned model is used to predict house prices. Using Python's Flask module, the front end is then merged with the trained model
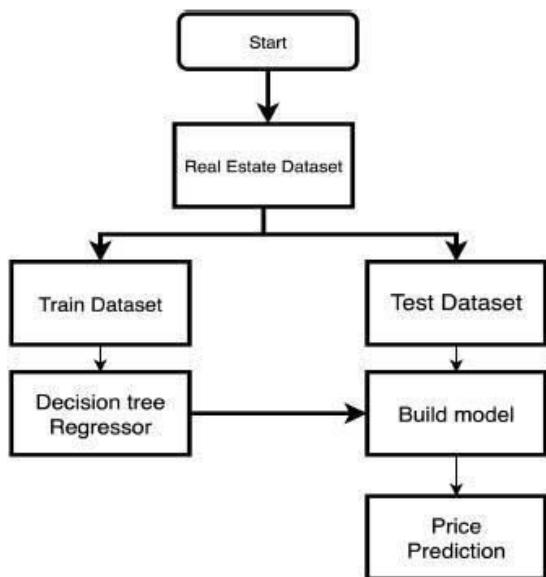
Fig 1. The generic flow of development

III. METHODOLOGY

A. STUDIEDALGORITHMS:

During the construction of this model, various regression algorithms were investigated.
SVM, Random Forest, Linear regression, Multiple linear regression, Decision Tree Regressor, and KNN were all evaluated on the training dataset. The decision tree regressor, on the other hand, forecasted house prices with the greatest precision. The algorithm selection is strongly influenced by the measurements and shape of data in the data used. The decision tree algorithm performed better on our dataset.

B. DECISION TREE REGRESSOR:

The decision tree regressor observes attributes and trains a model in the shape of a tree to forecast future data and generate realistic results. The max depth and min depth of a graph are learned by the

decision tree regressor, which then analyses the data according to the system.

Grid Search CV is a parameter tuning technique that allows you to easily generate and test a model for any set of measurement parameters defined in a grid. In this algorithm, Grid Search CV is used to evaluate the best value for max-depth, which is then used to construct the decisiontree.
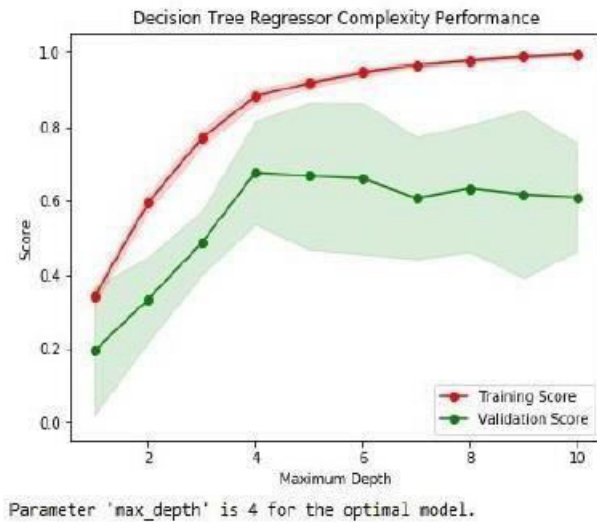
B. FLASK INTEGRATION
Following the efficient construction and delivery of the model, the next step is to combine it with the UI, which is done withflask.
Flask is a free and open-source web framework. Flask provides you with the tools, databases, and technology you'll need to build a web application. Flask is a Python model integration framework that makes it easy to linkroutes.
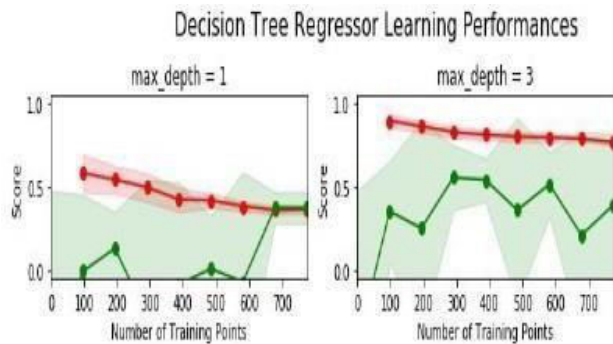
IV. IMPLEMENTATION

A. Datapreprocessig

B. The missing values for the age and floor parameters were dealt with.As a result, the goal attribute is not included in the training data list. The Pandas library is used for this. For statistical visualisation of the dataset, the min, max, standard deviation, and mean of the target attribute were calculated.
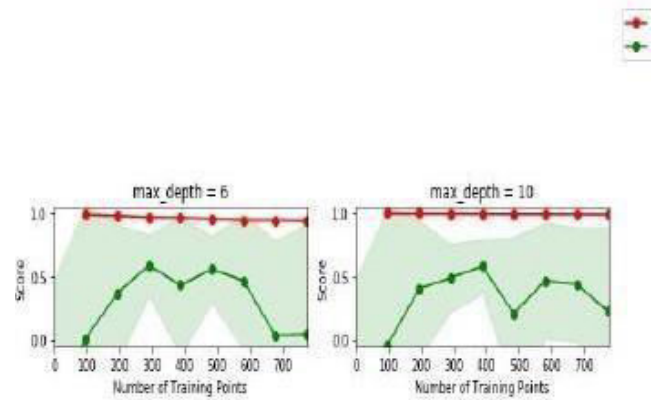
Parameter 'max_depth' is 4 for the optimal model.

Fig 3.Testing max-depth values(2) (On axis: Number of training pts. vs Score)

C. Max-depth:

Grid search cv, as previously reported, aids in the discovery of the tree's maximum depth. To see the different maximum depths and difficulty efficiency, we used Matplotlib. The below are the visualisations:

Fig 4. Max-depth value for optimal model

D. Fitting themodel:

The model is trained using a Decision tree regressor from the Scikit-learn library. The predict function is used to predict the test set outcomes.

Fig 2.Testing max-depth values(1) (On axis: Number of training pts. vs Score)

V. RESULTS:

The following shows the plot of predicted vs actual prices with the accuracy of prediction:
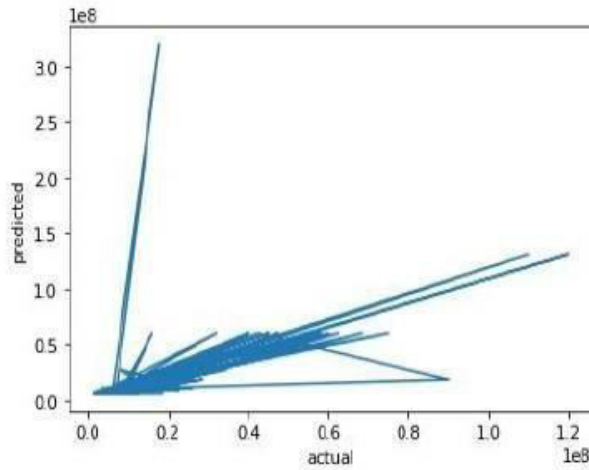
Fig 5. Actual vs predicted price graph based onthe dataset

Accuracy is nothing but the r2 scoreof the regression model.

VI. FUTURESCOPE

We hope to view a comparison analysis of the system's expected expense and the cost from real estate websites such as Housing.com with the same consumer feedback in the future. We would also propose genuine bequest assets to the client depending on the expected expense to rearrange it for the client. The existing dataset only includes cities in Mumbai; expanding it to other cities and states in India is the long-term goal. To make the system much more informative and user- friendly, we'll add Gmap. Within a 1 km radius of the given site, this will show local facilities such as hospitals and schools.

In this article, the Decision tree machine learning algorithm is used to construct a prediction model for potential real estate property selling values. Additional attributes such as air quality and crime rate were added to the dataset to help predict prices even higher. These features are not commonly found in other forecast application datasets, which distinguishes this method.

These factors influence people's decisions when purchasing a home, so why not factor them in when estimating house costs? The Flask Framework is used to combine the prepared show with the Client Interface. When it comes to estimating the costs of genuine bequests, the system has an accuracy of 89 percent

VIII. REFERENCES

L. Gatys, A. Ecker, and M. Bethge. A neural algorithm of artistic style. Nature Communications, 2015.

X. Glorot, A. Bordes, and Y. Bengio. Deep sparse rectifier neural networks. In AISTATS, 2011.

I. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio. Maxout networks.

S. Gross and M. Wilber. Training and investigating residual nets, 2016.

B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik.

VII.   CONCLUSION