Predicting house price using machine learning in ADC

ABSTRACT

In our ecosystem, real estate is clearly a distinct industry. Predicting house prices, significant housing characteristics, and many other things is made a lot easier by the capacity to extract data from raw data and extract essential information. Daily fluctuations in housing costs are still present, and they occasionally rise without regard to calculations. According to research, changes in property prices frequently have an impact on both homeowners and the real estate market. To analyze the key elements and the best predictive models for home prices, literature research is conducted. The analyses' findings supported the usage of artificial neural networks, support vector regression, and linear regression as the most effective modeling techniques. Our results also imply that real estate agents and geography play important roles in determining property prices. Finding the most crucial factors affecting housing prices and identifying the best machine learning model to utilize for this research would both be greatly aided by this study, especially for housing developers and researchers.

INTRODUCTION

In this report, we propose our system "House price prediction using Machine Learning". Along with other fundamental requirements like food, water, and many other things, a place to call house is one of a person's most basic wants. In the real estate sector, predicting house prices is essential to work since it aids buyers and sellers in making wise choices [1,2]. Numerous algorithms have been created to accurately anticipate property prices thanks to advances in machine learning. In this research, we use a dataset of real estate properties along with XGBoost, an advanced gradient boosting technique, to forecast house values [3 -5]. Powerful algorithm XGBoost effectively manages structured datasets [6-8].

PROBLEM STATEMENT

The asking price and general description are frequently presented independently from the generic and standardized real estate attributes. These qualities may be easily compared across the entire spectrum of potential houses because they are given separately and in a systematic manner. House sellers might list a summary of all the key aspects of the house in the description because every house also has distinctive elements, such as a particular view or style of washbasin. Potential purchasers can take into account all provided real estate features, but owing to the great diversity, it is almost not possible to provide an automatic comparison of all variables. This also applies in the opposite direction: house

sellers must evaluate the worth based on the attributes of the house in relation to the current market price of comparable houses. It is difficult to determine a fair market price due to the variety of features.

LITERATURE REVIEW

1. Sushant Kulkarni. (2021)Testing the dataset using four distinct retrogression algorithms—Velicet Lasso Regression, Logistic Retrogression, Decision Tree, and Support Vector Regression—is one of the approaches suggested in the study by Neelam Shinde and Kiran Gawande.

2. To predict the cost of resale homes, P. Durganjali suggested using classification algorithms. The selling price of a property is predicted in this study using a variety of classification methods, including Linear regression, Decision Tree, K-Means, and Random Forest.

3. Bengaluru has been chosen by Manasa and Gupta as the case study city. The square footage of the property, its location, and its amenities are all significant determinants of price. There are 9 different qualities employed.

4. According to Panjali and Vani, especially for those who plan to live there for a long time before selling it again. It also applies to people who want no risks taken when building their houses.

SYSTEM DESIGN AND ARCHITEC-TURE

1.Collection of Data In this phase, relevant data pertaining to house prices is gathered from reliable sources such as real estate websites and public datasets. The data may include features such as location, size, number of rooms, area_type, availability, and sale prices.

2. Data Pre-processing This phase involves cleaning and preparing the collected data for model training. Tasks such as handling missing values, removing outliers, normalizing numerical features, and encoding categorical variables are performed. Feature selection techniques can be applied to identify the most relevant attributes for predicting house prices.

3. Training the Model In this phase, various machine learning algorithms are applied to train a predictive model using the pre-processed data. Common approaches include linear regression, decision trees, random forests, or more advanced techniques like gradient boosting or neural networks.

4. Testing the Model Once the model is trained, it is evaluated using the testing dataset to assess its predictive capabilities. The model's performance is measured by comparing its predictions with the actual house prices in the testing set.

METHODOLOGY

To estimate housing values in this study, we used a number of well-known machine learning methods. Support vector machines (SVM), random forest, XGBoost, Lasso regression, and linear regression were some of the methods used in our investigation.

Algorithms: In the process of developing this model, various machine learning algorithms were studied. The model is trained on Support vector machines (SVM), random forest, XGBoost, Lasso regression, and linear regression.

IMPLEMENTATION

Here are the steps that we followed in implementation.

1. Data Collection Gather a dataset either from github or it will be also available on Kaggle, that includes relevant features of houses such as location, number of rooms, square feets, and sale prices.

2. Data Pre-processing Clean and prepare the collected data for model training. Handle missing values, perform feature scaling to bring features to a similar range, encode categorical variables, and address outliers.

3. Model Selection Choose a suitable machine learning algorithm for house price prediction, considering factors such as the dataset size, feature complexity, and interpretability requirements. Common algorithms include linear regression, decision trees, random forests, or more advanced techniques.

4. Exploratory Data Analysis In the exploratory data analysis (EDA) conducted for the house price prediction project, an image was generated to visualize the relationship between the variables "balcony," "bath," and "price."

5. Correlation Heatmap In our exploratory data analysis (EDA) for house price prediction, we created a correlation heatmap to examine the relationships between the variables bath, balcony, and price. The correlation heatmap visually represents the strength and direction of correlations between these variables.

RESULTS AND ANALYSIS

To use various machine learning algorithms for solving this problem. Random Forest achieves a high accuracy score of 0.903 and a low root mean squared error (RMSE) value of 44.032. This suggests that the Random Forest model captures the underlying patterns and relationships in the data effectively, resulting in accurate predictions of house prices.

CONCLUSION

The goal of the project "House Price Prediction Using Machine Learning" is to forecast house prices based on various features in the provided data. Our best accuracy was around 90% after we trained and tested the model. To make this model distinct from other prediction systems, we must include more parameters like tax and air quality.