PREDICTING HOUSE PRICES USING MACHINE LEARNING

By
Abish
Amaresh
Gopi Krishna
Altrin
Ahilesh

INTRODUCTION :

The aim of house price prediction is to create a model that can precisely estimate the price of a new house based on its attributes using previous data on house features (such as square footage, number of bedrooms and bathrooms, location, etc.) and their corresponding prices. In this project, we have applied these five algorithms namely linear regression, support vector machine, Lasso regression, Random Forest and XGBoost to predict house prices using a dataset of real estate properties. Because it can handle a large number of characteristics and capture intricate correlations between the features and the target variable (price), XGBoost is an effective algorithm for this purpose

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. In addition to outlining the property's essential features and capturing the

reader's curiosity, the house description functions as a persuasive tool. Housing prices are a significant indicator of the health of the economy, and both buyers and sellers are keenly interested in price points. In this study, explanatory variables that encompass a wide range of residential dwelling characteristics will be used to forecast house values. The objective of this project is to develop a regression model that can precisely calculate the house's price given its attributes.

Phases of Development :

Phase 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. Care should be taken to ensure the data is diverse and representative of the target market.

Phase 2 : Data Preprocessing

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. Additionally, data splitting techniques such as stratified sampling can be used to create training and testing datasets.

Phase 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. The training process involves fitting the

model to the training data, optimizing hyperparameters, and evaluating the model's performance using appropriate metrics such as mean squared error or R-squared.

Phase 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. Evaluation metrics such as mean absolute error or root mean squared error can be used to quantify the accuracy of the predictions.