PREDICTING HOUSE PRICES USING MACHINE LEARNING

ABSTRACT:

Machine Learning plays a virtual role from past years in normal speech command, product recommendation as well as in medical field also. Instead of this it provides better customer services and safer automobile system. This all of things shows that ML is trending technology in almost all fields so we are trying to coined up ML in our project. Nowadays the real estate market is a standout amongst the most focused regarding pricing and keep fluctuating. People are looking to buy a new home with their budgets and by analysing market strategies. But main disadvantage of current system is to calculate a price of house without necessary prediction about future market trends and result is price increase. So, the main aim of our project is to predict accurate price of house without any loss. There are many factors that have to be taken into consideration for predicting house price and try to predict efficient house pricing for customers with respect to their budget as well as also according to their priorities. So, we are creating a housing cost prediction model. By using Machine learning algorithms like Linear Regression, Decision Tree Regression, K-Means Regression and Random Forest Regression. This model will help people to put resources into a bequest without moving towards a broker. The result of this research provide that the Random Forest Regression gives maximum accuracy.

Keywords: Random forest regression, machine learning, datasets,predictions.

DESIGN THINKING:

Data Source:

We have gathered information about real estate from a variety of online real estate databases. We also gathered data from website such as kaggle, datahub.io etc... We need to gather accurate statistics categorized and organized. when we begin to address .Any machine learning issue must start with data require. The dataset must be valid or there will reason to analyze the data.

Dataset link: https://www.kaggle.com/datasets/vedavyasv/usa-housing

Data Preprocessing:

Our data is cleaned up at this step. Our dataset may contain values that are missing. Three methods exist to complete our missing values:

- Remove the data points that are missing.
- Remove the entire attribute.
- Set the value to a specific value, such as 0 or the median.

Feature selection:

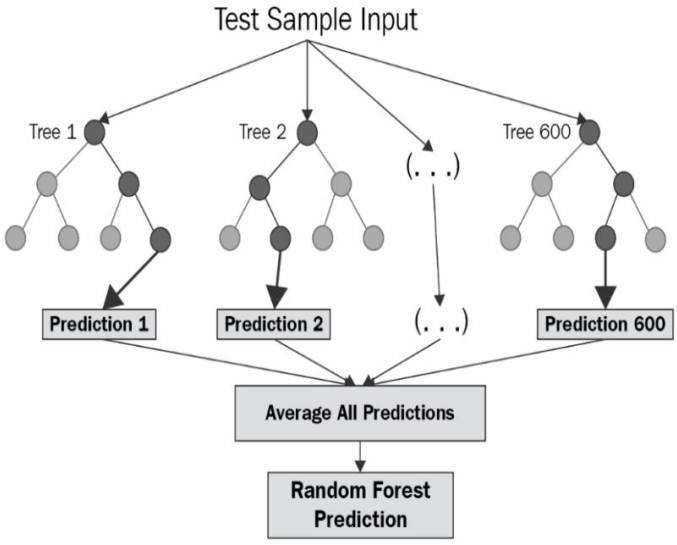
Feature selection is a crucial step in building a machine learning model for house price prediction. It involves selecting the most relevant and informative features (also known as variables or attributes) from your dataset while discarding irrelevant or redundant ones. Effective feature selection can improve model performance, reduce overfitting, and make your model more interpretable.

Create new features if they are likely to provide additional information. For example, you can calculate the price per square foot or create categorical variables from numeric ones

Model Selection:

Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as bagging .Random Forest has multiple decision trees as base learning models. We randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model. This part is called Bootstrap.

RandomForestRegressor – This is the regression model that is based upon the Random Forest model or the ensemble learning that we will be using in this article using the sklearn library.
Matplotlib/Seaborn – This library is used to draw visualizations.

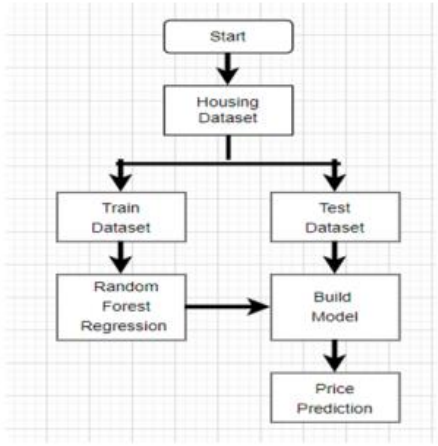


Random Forest Structure

Model Training:

Data is divided into two parts during this phase: training and testing. 80% of the data is used for training, and the remaining 20% is used for testing. Target variable is included in the training set. Different machine learning techniques are used to train the model while obtaining the

outcome. The results of these Random Forest Regressions are more accurate.



In the process of developing this model, various machine learning algorithms were studied. The model is trained on Leaner regression, Decision tree, K-mean and Random forest algorithms. Out of this Random Forest give a highest accuracy in prediction of housing prices. The decision to choose the algorithm is depends on the dimensions and type of data is used. Random Forest is best fit for our model.

Evaluation:

The machine learning models have been put to the test using data with known property prices in order to compare them. As a result, the data set is divided into training and testing sets. As the name implies, training data is utilized to develop the machine learning model and makes up a bigger portion of the split data set. The sales price of the properties is known to the model when it is fitted with the training data, allowing it to learn from the data. To gauge how successfully the machine learning model forecasts home prices, testing data is used. The test data is provided to the machine learning model without the pricing the model.

Conclusion :

Thus, the machine learning model using linear regression algorithm is very helpful in predicting the house prices for real estate customers.

Here we have used a supervised learning approach in machine learning field which will yield us a best possible result. The linear regression algorithm is used for this project because it is very simple to implement and hence gives accurate prediction of house price. Here in this project, we used python programming language. We also used different python packages like NumPy, pandas, matplotlib etc. For importing the dataset, and also for doing data pre-processing we used pandas. For doing exploratory data analysis we used matplotlib package in python.