PROJECT TITLE: House Price Prediction Using Machine Learning Project based on Python

ABSTRACT:

We all have experienced a time when we have to look up for a new house to buy. But then the journey begins with a lot of frauds, negotiating deals, researching the local areas and so on. Predicting house prices using machine learning involves training a model on a dataset that includes features like square footage, number of bedrooms, location, etc., and their corresponding sale prices.

INTRODUCTION:

The aim of this project was to develop a machine learning model capable of accurately predicting house prices based on a set of features such as square footage, number of bedrooms, location, etc. This report outlines the steps taken, the methodology employed, and the results obtained.The aim of this report is to outline the problem-solving approach used to predict house prices utilizing a machine learning model. This task is essential for various stakeholders in the real estate market, including buyers, sellers, and investors. To deal with this kind of issues Today we will be preparing a MACHINE LEARNING Based model, trained on the House Price Prediction Dataset.

DATASET INCLUDES:

The dataset contains 13 features :

| 1	Id	To count the records.
2	MSSubClass	Identifies the type of dwelling involved in the sale.
3	MSZoning	Identifies the general zoning classification of the sale.
4	LotArea	Lot size in square feet.
5	LotConfig	Configuration of the lot
6	BldgType	Type of dwelling
7	OverallCond	Rates the overall condition of the house
8	YearBuilt	Original construction year
9	YearRemodAdd	Remodel date (same as construction date if no remodeling or additions).
10	Exterior1st	Exterior covering on house
11	BsmtFinSF2	Type 2 finished square feet.
12	TotalBsmtSF	Total square feet of basement area
13	SalePrice	To be predicted

Importing Libraries and Dataset

Here we are using

Pandas – To load the Dataframe

Matplotlib – To visualize the data features i.e. barplot

Seaborn – To see the correlation between features using heatmap
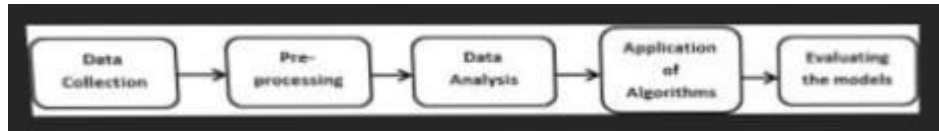
SVM – Support vector Machine

SVM can be used for both regression and classification model. It finds the hyperplane in the n-dimensional plane.

Random Forest Regression

Random Forest is an ensemble technique that uses multiple of decision trees and can be used for both regression and classification tasks.

Linear Regression

Linear Regression predicts the final output-dependent value based on the given independent features. Like, here we have to predict SalePrice depending on features like MSSubClass, YearBuilt, BldgType, Exterior1st etc

CatBoost Classifier

CatBoost is a machine learning algorithm implemented by Yandex and is open-source. It is simple to interface with deep learning frameworks such as Apple's Core ML and Google's TensorFlow. Performance, ease-of-use, and robustness are the main advantages of the CatBoost library.
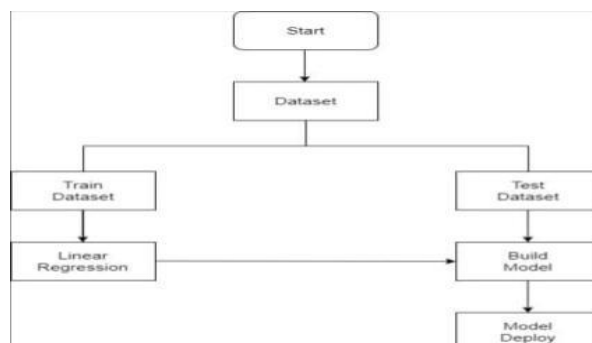
- People looking to buy a new home tend to be more conservative with their budgets and market strategies.
- This project aims to analyse various parameters like average income, average area etc. and predict the house price accordingly.
- This application will help customers to invest in an estate without approaching an agent
- To provide a better and fast way of performing operations.
- To provide proper house price to the customers.
- To eliminate need of real estate agent to gain information regarding house prices.
- To provide best price to user without getting cheated.
- To enable user to search home as per the budget.
- The aim is to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. By analyzing previous market trends and price ranges, and also upcoming developments future prices will be predicted.
- House prices increase every year, so there is a need for a system to predict house prices in the future.
- House price prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house.
- We use linear regression algorithm in machine learning for predicting the house price trends

www.economicshelp.org

PROPOSED SYSTEM PHASES

Phase 1: Collection of data

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.

Phase 2: Data preprocessing

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.

Phase 3: Training the model

Since the data is broken down into two modules: a Training set and Test set, we must initially train the model. The training set includes the target variable. The decision tree regressor algorithm is applied to the training data set. The Decision tree builds a regression model in the form of a tree structure.

Phase 4: Testing and Integrating with UI

The trained model is applied to test dataset and house prices are predicted. The trained model is then integrated with the front end using Flask in python