PREDICTING HOUSE PRICE USING MACHINE LEARNING

ABSTRACT:

This project provides us an overview on how to predict house prices using various machine learning models with the help of different python libraries. This proposed model considers as the most accurate model used for calculating the house price and provides a most accurate prediction. This provides a brief introduction which will be needed to predict the house price. This project consists of what and how the house price model works with the assistance of machine learning technique using scikit-learn and which datasets we will be using in our proposed model. Predicting the price of a house helps for determine the selling price of the house in a particular region and it help people to find the correct time to buy a home. In this task on House Price Prediction using machine learning, our task is to use data to create a machine learning model to predict house prices in the given region. We will implement a linear regression algorithm on our dataset. By using real world data entities, we are going to predict the price of the house in that area. For better results we require data pre-processing units to improve the efficiency of the model. for this project we are using supervised learning, which is a part of machine learning. We have to go through different attributes of the dataset. Keywords: - Python, Machine learning, scikit-learn, python libraries, data pre-processing, Linear regression algorithm, Supervised learning.

INTRODUCTION:

One of the basic requirements of livelihood in the recent world is to buy a house of your own. The price of the house may depend on various factors. Real estate agents and many who are involved in selling the house want a price tag on the house which would be the real worth of buying the house. The prediction of the price of the house is often very hard for the inexperienced.

House is a fundamental need to someone and them costs range from place to place primarily based totally at the facilities to be had like parking area, location, etc. The residence price is a factor that issues a lot of citizens either wealthy or white-collar magnificence as possible by no means decide or gauge the value of a residence primarily based totally on place or places of work accessible. Purchasing a residence is the best and unique desire of a own circle of relatives because this expends the whole thing of their funding budget and every now and then cover them under loan. It is a hard challenge for expecting the correct value of residence price. This proposed version could make viable who are expecting the precise costs of house. This model of Linear regression in machine learning takes the internal factor of house valuation dependencies like area, no of bedrooms, locality etc. and external factors like air pollution and crime rates. This Linear regression in machine learning gives the output of price of the house with more accuracy.

Domain problem to ML problem: -

Domain Problem: In this project we are going to predict the price of a house for real estate customers based on their preferences like no of bedrooms, area of the house, locality etc. ML problem:

This problem can be solved using supervised learning in ML. It is treated as a Regression problem as the target is continuously valued.

Data collection and preparation: -

This phase consists of three parts they are: -

1.Data collection

2.Data pre-processing

3.EDA and feature Engineering Data collection:-In order to proceed further into this project, first of all we should collect the data, after collecting a dataset, then we should pre-process the data and then we do the exploratory data analysis on the given data set.

Data **pre**-processing:-

This pre-processing of data consists of Handling missing values which means some of values in the dataset may be missing. We should handle those missing values. Handling duplicates which mean some of the values in the dataset may be their multiple times we should handle those duplicate values that means we should delete those duplicates. Handling outliers, it consists of identifying the outliers and hence removing those outliers.

EDA and feature Engineering:-

By using matplotlib package in python we will do EDA on a given dataset. Normally we will do EDA on a given data set to establish a relation between given feature of the data set.

Linear Regression Algorithm

Linear regression is a supervised learning algorithm in machine learning field. It is a statistical method which aims for developing a linear relation between a dependent variable or target variable and one or more independent variables by developing a linear equation to the given data. Formulae: Y=b0+b1x+E Where, Y=target variable or dependent variable X=independent variable E=error b1=linear regression coefficient b0=intercept

Correlation: -

It is a statistical tool that describes the relationship existing between any two variables. The strength of a correlation is of two types they are: Strong correlation Weak correlation If there exists a weak correlation then we are not able to make accurate predictions. If there exists a strong correlation, we can make accurate predictions.

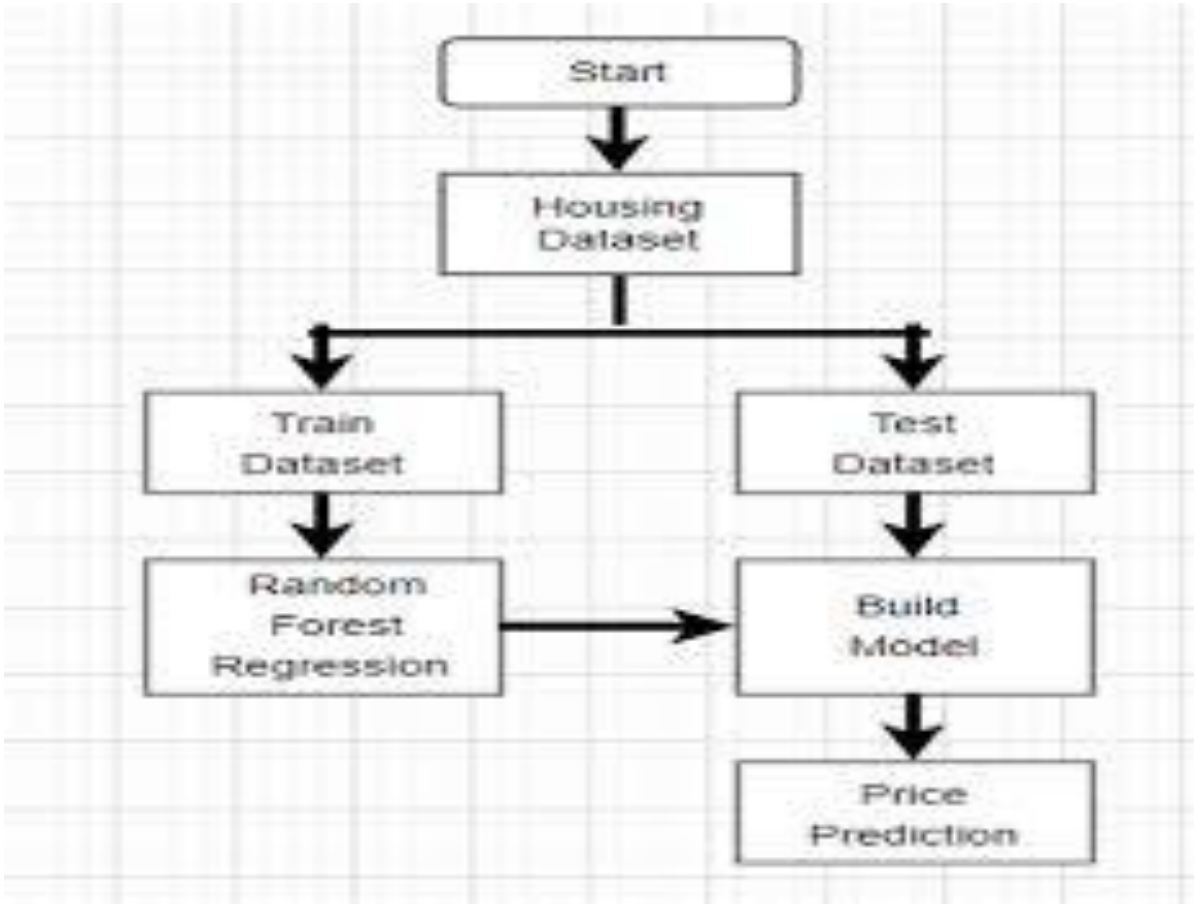
Confidence: -

If we have more data, then we have more confidence on the predictions made based on the relationship between the variables.

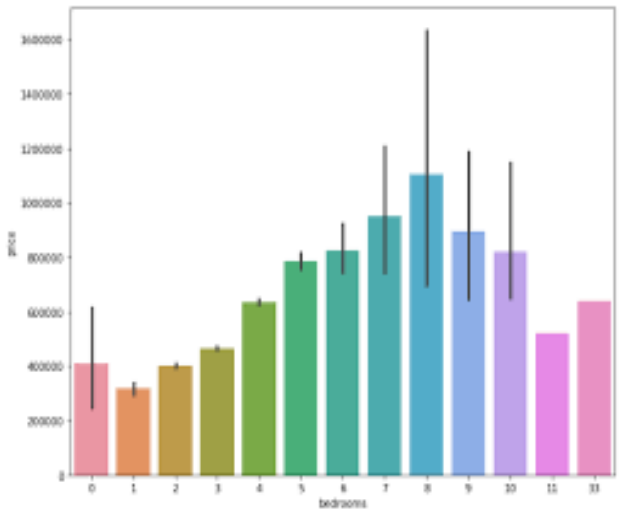
Correlation matrix: -

It is a table that contains the correlation coefficients of the features.it is normally used to identify the useful features and hence drops the useless ones.

Flow chart



Data graph:



Actually, Linear regression are of two types Simple linear regression Multiple linear regression

Simple linear regression:

If a regression algorithm contains a single independent variable that is used to predict a target variable, then such an algorithm is called simple linear regression.

Multiple linear regression: -

If a regression algorithm contains more than one independent variable that is used to predict the target variable, then such an algorithm is called multiple linear regression. Before diving deep into linear regression algorithm, we should know some of the key statistical concepts like.

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.