Predicting house prices using machine learning

INNOVATION:

GRADIENT BOOSTING FOR IMPROVED

PREDICTING ACCRACY

INTRODUCTION

Comparing the housing crisis of the last decade with the 1930s-great depression, Global Research estimated that with the current 316 million Americans, the foreclosure rate was higher than the great depression era. During the bubbling phase of the housing market, real estate market saw a large increase in transactions. Buyers were willing to make offers to available properties and home sellers made the most of the situation. Mortgage lenders developed new models of lending like stated income and no credit verification.

SYSTEM ARCHITECTURE:

The first step in this is collection of raw data from the various sources and dataset can be of any historical data of the organization. From the raw data we can extract the attributes which are used for the prediction. After extraction, we can train the data model using these previous datasets. Here we should give Testing data as input to data analytical tool.

ALGORITHM:
Gradient Boosting Regressor:

ALGORITHM: GRADIENT BOOST REGRESSOR: We used the python pandas library for data processing which combined different datasets into a data frame. The raw data makes us to prepare the data for feature identification. Gradient Boosting for regression builds an additive model in a forward stage wise fashion. It allows for the optimization of arbitrary differentiable loss functions.

A weak hypothesis is defined as one whose performance is at least slightly better than random chance.

Step 1: Load the data set df = pd.read_csv("ml_house_data_set.csv"

Step 2: Replace categorical data with one-hot encoded data.

Step 3: Remove the sale price from the feature data .

Step 4: Create the features and labels X and Y arrays.

Step 5: Split the data set in a training set (70%) and a test set (30%).

Step 6: Fit regression model.

Step 7: Save the trained model to a file trained_house_classifier_model.pkl

Step 8: Predict house worth using predict function

MODULE:

I. Data Collection

Firstly, Dataset can be collected from various sources of any organization. The right dataset helps for the prediction and it can be manipulated as per our requirement. Our data mainly consists of the attributes of houses available in particular Area. The data can be collected from the organization based on the house area, no. of bed rooms, bath rooms, availability of swimming pool, fire place. By collecting these it makes accurate in prediction.

II. Data Processing

At the beginning, when the data was collected, all the values of the attributes selected were continuous numeric values. Data transformation was applied by generalizing data to a higher-level concept so as all the values became discrete. The criterion that was made to transform the numeric values of each attribute to discrete values depended on the closing price of the house. The attribute values of the houses of area are taken to predict the price of the house in that area.

III. Training the Data

After the data has been prepared and transformed, the next step was to build the classification model using the decision tree technique. The decision tree technique was selected because the construction of decision tree classifiers does not require any domain knowledge, we can done by using the Decision Tree Classifier () in which 70 % of the data is used for training the data and another 30 % is used for testing the data.

IV. Deploying the Model

The classification rules are generated from the decision tree algorithm. The trained data can be used for the Testing the data. It helps to give the output or accurate Predicted price of the stock using this model.

SUBMITTED BY,
BNS PROJECT TEAM.