Abstract:

House Price Index (HPI) is commonly used to estimate the changes in housing price. Since housing price is strongly correlated to other factors such as location, area, population, it requires other information apart from HPI to predict individual housing price.
There has been a considerably large number of papers adopting traditional machine learning approaches to predict housing prices accurately, but they rarely concern about the performance of individual models and neglect the less popular yet complex models.
As a result, to explore various impacts of features on prediction methods, this paper will apply both traditional and advanced machine learning approaches to investigate the difference among several advanced models. This paper will also comprehensively validate multiple techniques in model implementation on regression and provide an optimistic result for housing price prediction.

Chapter-1

Introduction:

House is one of human life's most essential needs, along with other fundamental needs such as food, water, and much more. Demand for houses grew rapidly over the years as people's living standards improved. While there are people who make their house as an investment and property, yet most people around the world are buying a house as their shelter or as their livelihood. According to housing markets have a positive impact on a country's currency, which is an important national economy scale. Homeowners will purchase goods such as furniture and household equipment for their home, and homebuilders or contractors will purchase raw material to build houses to satisfy house demand, which is an indication of the economic wave effect created by the new house supply. Besides that, consumers have capital to make a large investment, and the construction industry is in good condition can be seen through a country's high level of house supply.

In a very general way, recommender systems are algorithms aimed at suggesting relevant items to users (items being movies to watch, text to read, products to buy or anything else depending on industries). This website also provides an option for recommendations. The type of recommendation system is content based recommendation. In this project, we are using two datasets using the concept of web scraping. One dataset consists of some features such as location, BHK, floor, furnished etc. with different cities in Mumbai. This dataset is used for prediction. The other dataset consists of the House Price index of Mumbai for the last 10 years. This
dataset is used for forecasting.

An increase in house demand occurs each year, indirectly causing house price increases every year. The problem arises when there are numerous variables such as location and property demand that may influence the house price, thus most stakeholders including buyers and developers, house builders and the real estate industry would like to know the exact attributes or the accurate factors influencing the house price to help investors make decisions and help house builders set the house price. House price prediction can be done by using a multiple prediction models (Machine Learning Model) such as support vector regression, artificial neural network, and more. There are many benefits that home buyers, property investors, and house builders can reap from the house-price model. This model will provide a lot of information and knowledge to home buyers, property investors and house builders, such as the valuation of house prices in the present market, which will help them determine house prices. Meanwhile, this model can help potential buyers decide the characteristics of a house they want according to their budget. Previous studies focused on analyzing the attributes that affect house price and predicting house price based on the model of machine learning separately. However, this article combines such a both predicting house price and attributes together.

Chapter 2

Literature review

2.1. HOUSE PRICE PREDICTION FORECASTING AND RECOMMENDATION SYSTEM USING MACHINE LEARNING

Abstract-
The relationship between house prices and the economy is an important motivating factor for predicting house prices. A property's value is important in real estate transactions. Housing price trends are not only the concern of buyers and sellers, but it also indicates the current economic situation. Therefore, it is important to predict housing prices without bias to help both the buyers and sellers make their decisions. In this project,

we are going to create a website where user have to add some property details for predicting the house price, enter date for forecasting the price till that
date and budget range for recommending best location. This project uses two datasets, one includes some features and large entries of housing sales in Mumbai and another contains house price index of Mumbai. We are using different feature selection methods and feature extraction method with Multiple Linear Regression to predict the current house price and using ARIMA model for forecasting the price after few years in Mumbai and also uses content based recommendation system to recommend best location according to their budget in nearby area of interest.

INTRODUCTION

Investment is a business activity on which most people are interested in this globalization era. There are several objects that are often used for investment, for
example, gold, stocks and property. In particular, property investment has increased significantly. Housing price trends are not only the concern of buyers and sellers, but it also indicates the current economic situation. There are many factors which has
impact on house prices, such as location, BHK, floor etc. Also, a location with a great accessibility to highways, expressways, schools, shopping malls and local employment opportunities contributes to the rise in house price. Manual house prediction becomes difficult, hence there are many systems developed for
house price prediction. The aim of this system is to create a website through which the user can give his house requirements as input which is then passed on to the linear regression model for predicting the house price. The website also allows user to forecast the predicted house price to a particular date which is also specified by the user. This is done by using another model known as the ARIMA(Auto Regressive Integrated Moving Average Model). During the last few decades, with the rise of Youtube, Amazon, Netflix and many other such web services, recommender systems have taken more and more place in our lives. From e-commerce (suggest to buyers
articles that could interest them) to online advertisement (suggest to users the right contents, matching their preferences), recommender systems are today unavoidable in our daily online journeys. In a very general way, recommender systems are algorithms aimed at suggesting relevant items to users (items being movies to watch, text to read, products to buy or anything else depending on industries). This website also provides an

option for recommendations. The type of recommendation system is content based
recommendation. In this project, we are using two datasets which are extracted from Makaan.com by using the concept of web scraping. One dataset consists of some features such as location, BHK, floor, furnished etc. with different cities in Mumbai. This dataset is used for prediction. The other dataset consists of the House
Price index of Mumbai for the last 10 years. This dataset is used for forecasting.

REQUIREMENT ANALYSIS

Requirement analysis gives a minimum requirement that a system should have to make the software to work properly. This application can work on any website. Usually the requirement specification will be the same as that of the operating system.

A. Functional Requirements:

FR1: USER INTERFACE:
The user interface will be a website. The user has to enter all the attributes
correctly and in the required format.

FR2: PROPER FORECASTING:
The system has to properly predict the price of the house according to the input given by the user.

FR3: RECOMMENDATION SYSTEM:
According to the input given by the user, the recommendation system will recommend the best property.

FR4: DATABASE:
Dataset should contain large number of entities so that it will increase the accuracy of the predicted price and suggest a better property

B. Non Functional Requirements:

QR1: Platform Independent:
The application would be platform independent if all the requirements are installed in the device.

QR2: Performance:
The application should have better accuracy and should provide the information in less time.

QR3: Capacity:
The capacity of the storage should be high so that large amount of data can be stored in order to train the model.

C. Software Requirements:

1. Coding Language: Python3, HTML, Python Flask

2. Coding software : Anaconda, Spyder, Jupyter Notebook, Sublime text 3

Safety Requirements:

For every input given by user, no incorrect format of data can be given as an input to the system which can be of various forms. All the data fields must be filled by
the user to get the Output. The date provided for forecasting should be given of the future not that of the past.

DESIGN AND IMPLEMENTATION

User interface:

The user interface for our project is Website. For this software, the users are the businessmans, investors and other people searching for property. They have to
enter details about the property they want and then the software will give the accurate predicted value. User can also forecast the predicted value by entering date. In this application, the user have to enter information on website about the users location such as number of floors, area in sq. feet, location, bhk, furnishing, date for forecasting and budget.

DataSets:

• Dataset is Extracted from Makaan.com by using concept of Web Scraping for house price prediction purpose and downloaded another dataset from TradingEconomics for forecasting.
• Dataset used for prediction contains names of all cities in and nearby Mumbai with their BHK, Sq.ft, Furnished or not, Floor No. and Prices.
• It contains 160000 entries which contains 1400 different cities and places in mumbai.
• Dataset used for forecasting contains House price index according to date for year 2010 to 2019.

Data Preparation:

To prepare the dataset for the prediction system, some changes were made:
1. Binary categorical variables (furniture) is represented using one binary digit (i.e.
(Furnishing) 0 = Not Furnished, 1 = Furnished).
2. Also by using Label encoder names of places is to be converted into values as linear regression model is to be trained by using values.
3. As the price for properties are often quoted in Lakhs, we have rounded our dependent variable to the nearest thousand, which also helps with the numerical stability of the model.

Methodology:

Linear Regression:

• In this Project, we have used Linear Regression Algorithm for predicting the current house price.
• The Linear Regression Algorithm accepts two variables Independent variable (X) and Dependent variable (Y).
• We have used sklearn Library for importing Linear Regression model.
• The dataset containing different cities with their features and prices is used for training Linear Regression Model.
• The dataset entities will be divided into two parts 80% for training and 20% for testing.
• Linear Regression model will be trained using X_train Independent variable entries and Y_train Dependent variable entries.
• The trained model will be tested upon the 20% test dataset entities. After training and testing the model will be use for prediction purpose.
• The accuracy for trained linear regression model is 86.67%.

ARIMA(Auto Regressive Integrated Moving Average Model):
• ARIMA Model is widely used for Forecasting purpose like stock, temperature forecasting, sales predictions etc.
• In this project, the ARIMA is used to forecast house price for a particular date which is gives by the user.
• ARIMA Model is the combination for three methods for forecasting which are
AutoRegressive (AR) Model, Integrated differencing and Moving Average (MA) Model.

1. AutoRegressive(AR) Model:
Yt depends only on past values Yt-1, Yt-2, so on. Yt = F(Yt-1,Yt-2,Yt-3,…) if no.of past values (p) increases then the accuracy of the model increases.
2. Moving Average (MA) Model **:**
Yt depends only on past error terms. Yt=F(Et, Et-1, Et-2,…) the no.of past error
terms taken is mostly 0, 1 or 2. The No.of error terms is denoted by ‘q’.
3. **Integrated Differencing:**
* In ARIMA Model Series is need to be “**Strictly Stationary**” *
It means, mean, variance and covariance must be constant over the time period. If
series is not stationary then it is converted to stationary using differencing parameter
‘d’ which is generally equal to 1 or 2.
 The dataset entities will be divided into two parts 80% for training and 20% for testing.

The ARIMA model is imported from statsmodel library which takes training dataset and order of (p, d, q) as input. After training the Model will be use for forecasting purpose.
This project contains ARIMA model with 87% of accuracy.

Content Based Recommendation:

Recommendation system is a machine learning system that gives generalized recommendation to its users based on some data or using users preferences. It
produce a list of recommendations in any of the two ways:

• Collaborative filtering: Collaborative filtering approaches build a model from user's past behaviour (i.e. items purchased or searched by the user) as well as similar decisions made by other users.

• Content-based filtering: Content-based filtering approaches uses a series of discrete
characteristics of an item in order to recommend additional items with similar
properties.

• We are going to use content based filtering methods in our project.

• For example in our project data set contains tupples like place ,square feet , number of bhk , flat is furnished or not and floor number at which given flat is situated.

• Suppose client wants a new two bhk fully furnished flat in a area like diva of 300 sqrt feet at a floor maximum up to 5.

• After that recommendation system takes this and makes some assumptions client has
entered floor number is 5 than system will search the flate for floor number 4 to 6 from training dataset that means if product is not available then it will try to give maximum similar type of product.

• Our will match the every preference made by the user with the values present in the training data set and will try to give similar type of product.

Model Procedure:

The trained linear regression model is given the user entered property details as input and model will return predicted value which is pass from flask to website. For
Arima model the input will the user entered date. The ARIMA model will give House Price Index (HPI) as output which is converted to House price by using formula

Current House Price Value * Future HPI = Future House Price Value * Current HPI

Then the Forecasted house price is displayed on web by flask.

Connectivity:

- The website is connected to backend by using framework called python flask.
- The flask provides a local IP address through which the websites is connected.
- When user enter details about property on website, the IP address provided by flask is used to pass data to flask.
- In python flask program, the trained linear regression model is imported by using library joblib and property details fetched from URL is given to the trained model. The output is given as predicted price which is displayed on screen.
- Similarly, User entered date is also fetch from URL for forecasting. This date is given to the imported ARIMA model which gives forecasted HPI(House Price Index) as output.
- The forecasted HPI is used to calculate forecasted price which is return to the website.
- For Recommendation, the user entered budget range is used to filter out all the properties which satisfy user property requirement this is called as content based filtering.
- The filtered properties are sent to website in HTML table format.

TECHNOLOGY USED

Documentation Tools

● Microsoft Office Word.
● Snipping Tools (For Screenshots).
● StarUML (For UML Diagrams).
● LucidChart.
● Microsoft Excel

Language Used

The "**HOUSE PRICE FORECASTING AND RECOMMENDATION SYSTEM**" will be used to for predicting the house price, forecasting that price and also to get best recommendation according to users requirement. This application can be run using website. We are using Python3 for making machine learning model and Python flask for connectivity and HTML to develop our web page. We are using anaconda which contains a software Spyder and Jupyter Notebook. Spyder contains all updated and latest libraries of python which will be very useful for implementing machine learning model linear regression, ARIMA model and content based
Recommendation system. Sublime Text 3 will be used for implementing HTML web page which will be user interface.

RESULT

In this project frontend contains website . On that website first we have to give input likes Location, BHK, Sq.ft , Furnish and floor for prediction.

CONCLUSION

In this project, the website allows the user to give property details according to his/her requirement. The system makes optimal use of the Data mining Algorithm i.e Linear Regression, ARIMA Model along with Content Based Recommendation System. The
Linear Regression algorithm is used to predict the house price according to the property requirement given by the customer with accuracy of 86.7%. ARIMA Model is used for Forecasting the predicted house price with an accuracy of 87%. Content based Recommendation system will help the user to get the best and relevant real estates residential properties according to the budget given by the user. The
connectivity between website and models is done by using python flask. The main objective of using this prediction, forecasting and recommendation system is to reduce the human physical calculation, time and carry out the whole process at ease.

2.2. House Price Prediction using a Machine Learning Model: A Survey of Literature

Abstract:

Data mining is now commonly applied in the real estate market. Data mining's ability to extract relevant knowledge from raw data makes it very useful to predict house prices, key housing attributes, and many more. Research has stated that the fluctuations in house prices are often a concern for house owners and the real estate market. A survey of literature is carried out to analyze the relevant attributes and the most efficient models to forecast the house prices. The findings of this analysis verified the use of the Artificial Neural Network, Support Vector Regression and XGBoost as the most efficient models compared to others. Moreover, our findings also suggest that locational attributes and structural attributes are prominent factors in predicting house prices. This study will be of tremendous benefit, especially to housing developers and researchers, to ascertain the most significant attributes to determine house prices and to acknowledge the best machine learning model to be used to conduct a study in this field.

fitting problems, while ensuring a single optimum solution by minimizing structural risks and empirical risks. In this field of study, support vector regression is used to collect details on neighborhood, structural and locational attributes.

C. Artificial neural Network

In 1958, created artificial neural network known as ANN. Walter Pitts and Warren McCulloch published a paper entitled "A Logical Calculus of Ideas Immanent in Nervous Activity" in the year 1943 which notes that a neural network may artificially be created, based on the role and structure of a biological neural network. In another research, as this model would often promote learning, artificial neural networks are claimed to be artificial brain diagrams.

The artificial neural network model has always been selected when a non-linear attribute is involved. The analysis of home price estimation should also use this model as a spatial consideration for the price of housing is also non-linear. Therefore, as in, their study produces a good result, thus it is promising to provide an exact predictive model utilizing the artificial neural network algorithm. This system, however, has very limited performance. ANN can model complex non-linear relationships as house price predictions involve many non-linear variables.

D. Gradient Boost

Gradient boosting was created by in 1999 and is a commonly used machine learning algorithm because of its performance, consistency and interpretability. Gradient boosting delivers state-of-the-art in various machine learning activities, such as multistage classification, click prediction and ranking. With the advent of big data in recent years, gradient boosting faces new challenges, especially with regard to the balance between accuracy and performance. There are few parameters for gradient boosting. To ensure a dynamic balance between fit and regularity, the following

steps can be taken to select parameters: (1) Setting regularization parameters (lambda, alpha), (2) reducing learning rate and decide those optimal parameters again.

Finding and Discussion

The associations between the house price and predicting model included in this segment have been explored. In addition, the impact of various attributes on specific model have also been evaluated and debated. Based on reviewing numerous papers, there are several attributes used by researchers in their work to forecast

house prices. All of these attributes can be divided into 4 main categories which are locational, structural, neighborhood and economic attributes. The locational attribute consists of variables which described the