HOUSE PRICE PREDICTION FORECASTING AND RECOMMENDATION SYSTEM USING MACHINE LEARNING

Ashutosh Sharma[1], Pranav Sonawale[2], Deeksha Ghonasgi[3], Shreya Patankar[4]

[1-3]Student- Ashutosh Sharma, Dept. of computer Engineering, Datta Meghe college, Maharashtra, India
[4]Professor- Shreya Patankar, Dept. of computer Engineering, Datta Meghe college, Maharashtra, India

---***---

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.*

Keywords- House price prediction and forecasting using machine learning algorithm, Recommendation of house according to user choice.

1. 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.

2. LITERATURE REVIEW

The Real Estate has had a significant impact on all aspects of our society. As today society relies more and more on the Technology, the dependability of accurate prediction and recommending applications by using the technology has become increasingly important. To make these applications more dependable, for the past decade researchers have proposed various techniques to implement Machine Learning. Our literature search for related studies retrieved 3 papers in the area of

Machine Learning and AI, which have appeared between 2000 and 2013.

We are using Multiple Linear Regression Model for prediction and ARIMA Model for forecasting.

Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the linear relationship between the explanatory (independent) variables and response (dependent) variable.

In essence, multiple regression is the extension of ordinary least-squares (OLS) regression that involves more than one explanatory variable. The Formula for Multiple Linear Regression Is:

$yi=\beta0+\beta1xi1+\beta2xi2+...+\beta pxip+\epsilon$

where,

for i= n observations:

yi=dependent variable

xi = explanatory variables

β0=y-intercept (constant term)

βp=slope coefficients for each explanatory variable

ϵ=the model's error term (also known as the residuals)

The multiple regression model is based on the following assumptions:

There is a linear relationship between the dependent variables and the independent variables.

The independent variables are not too highly correlated with each other.

y_i observations are selected independently and randomly from the population.

Residuals should be normally distributed with a mean of 0 and variance *σ*.

The coefficient of determination (R-squared) is a statistical metric that is used to measure how much of the variation in outcome can be explained by the variation in the independent variables. R^2 always increases as more predictors are added to the MLR model even though the predictors may not be related to the outcome variable.

R^2 by itself can't thus be used to identify which predictors should be included in a model and which should be excluded. R^2 can only be between 0 and 1, where 0 indicates that the outcome cannot be predicted by any of the independent variables and 1 indicates that the outcome can be predicted without error from the independent variables.

When interpreting the results of a multiple regression, beta coefficients are valid while holding all other variables constant ("all else equal"). The output from a multiple regression can be displayed horizontally as an equation, or vertically in table form.

Recommender System is a system that seeks to predict or filter preferences according to the user's choices. Recommender systems are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general.

Recommender systems 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. This model is then used to predict items (or ratings for items) that user may have an interest in but this type of algorithm have some limitations.

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. Content-based filtering methods are totally based on a description of the item and a profile of the user's preferences. It recommends items based on user's past preferences. This is one of the metric that we can use when calculating similarity, between users or contents.

Cosine similarity

The dot product between two vectors is equal to the projection of one of them on the other. Therefore, the dot product between two identical vectors (i.e. with identical components) is equal to their squared module, while if the two are perpendicular (i.e. they do not share any directions), the dot product is zero. Generally, for *n*-dimensional vectors, the dot product can be calculated as shown below.

$$u \cdot v = [u_1\, u_2\, \ldots\, u_n] \cdot \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n = \sum_{i=1}^{n} u_i v_i$$

Dot product.

The dot product is important when defining the similarity, as it is directly connected to it. The definition of similarity between two vectors **u** and **v** is, in fact, the ratio between their dot product and the product of their magnitudes.

Similarity = cos(Θ) = $\frac{\sum_{i=1}^{n} u_i v_i}{\sqrt{\sum_{i=1}^{n} u_i^2}\sqrt{\sum_{i=1}^{n} v_i^2}}$

By applying the definition of similarity, this will be in fact equal to 1 if the two vectors are identical, and it will be 0 if the two are orthogonal. In other words, the similarity is a number bounded between 0 and 1 that tells us how much the two vectors are similar.

Content based filtering can also be achieved by using logical operators like AND, OR, NOT and comparison operators like < , >,= etc.

Consider the following table:

PLACE	BHK	Sq.ft	FLOOR	PRICE
Diva	1	350	6	3161000

We can recommend the best location to user for purchase by applying logical and comparison operators between data entered by user and data values present in the table. By using this operators we will exclude unwanted data from table and only will consider best suitable data for user.

e.g. if {(maximum budget price entered by the user) < (price values present in table for particular home)}

using this formula we can exclude home location which are not falling in budget range of customer.

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 short for 'Auto Regressive Integrated Moving Average' is actually a class of models that explains a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values.

An ARIMA model is characterized by 3 terms: p, d, q

where,

p is the order of the AR term

q is the order of the MA term

d is the number of differencing required to make the time series stationary.

Because, term 'Auto Regressive' in ARIMA means it is a linear regression model that uses its own lags as predictors. Linear regression models, as we know, work best when the predictors are not correlated and are independent of each other.

So how to make a series stationary?

The most common approach is to difference it. That is, subtract the previous value from the current value. Sometimes, depending on the complexity of the series, more than one differencing may be needed.

The value of d, therefore, is the minimum number of differencing needed to make the series stationary. And if the time series is already stationary, then d = 0.

'p' is the order of the 'Auto Regressive' (AR) term. It refers to the number of lags of Y to be used as predictors. And 'q' is the order of the 'Moving Average' (MA) term. It refers to the number of lagged forecast errors that should go into the ARIMA Model.

A pure Auto Regressive (AR only) model is one where Yt depends only on its own lags. That is, Yt is a function of the 'lags of Yt'.

$$y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} + \varepsilon_1$$

where, $Y{t-1}$ *is the lag1 of the series,* $\beta1$ is the coefficient of lag1 that the model estimates and α is the intercept term, also estimated by the model.

Likewise a pure Moving Average (MA only) model is one where Yt depends only on the lagged forecast errors.

$$y_t = \alpha + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + .. + \phi_q \varepsilon_{t-q}$$

where the error terms are the errors of the autoregressive models of the respective lags. The errors Et and E(t-1) are the errors from the following equations :

$$y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_0 Y_0 + \varepsilon_t$$

$$y_{t-1} = \beta_1 Y_{t-2} + \beta_2 Y_{t-3} + .. + \beta_0 Y_0 + \varepsilon_{t-1}$$

That was AR and MA models respectively.
An ARIMA model is one where the time series was differenced at least once to make it stationary and you combine the AR and the MA terms. So the equation becomes:

$$y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + .. + \phi_q \varepsilon_{t-q}$$

ARIMA model in words:

Predicted Yt = Constant + Linear combination Lags of Y (upto p lags) + Linear Combination of Lagged forecast errors (upto q lags)
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.

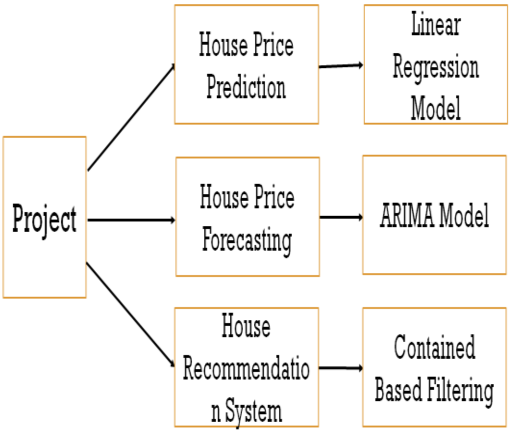


Fig -1: Algorithm used for various process

We firstly took a brief idea and knowledge about how Real estate and House price prediction and recommendation system actually works and then implemented Machine Learning to predict house prices based on certain criterias(no of floors, area in sq. feet, location, bhk, etc.).We used concepts of Machine Learning such as Linear Regression and Multiple Linear Regression to help us predict the prices along with that we used ARIMA model to forecast the house price after few years. We also used concept of content based filtering to recommend the best house for buyers.

3. 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.

4. DESIGN PROCESS

The diagram of the model is given below:

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.

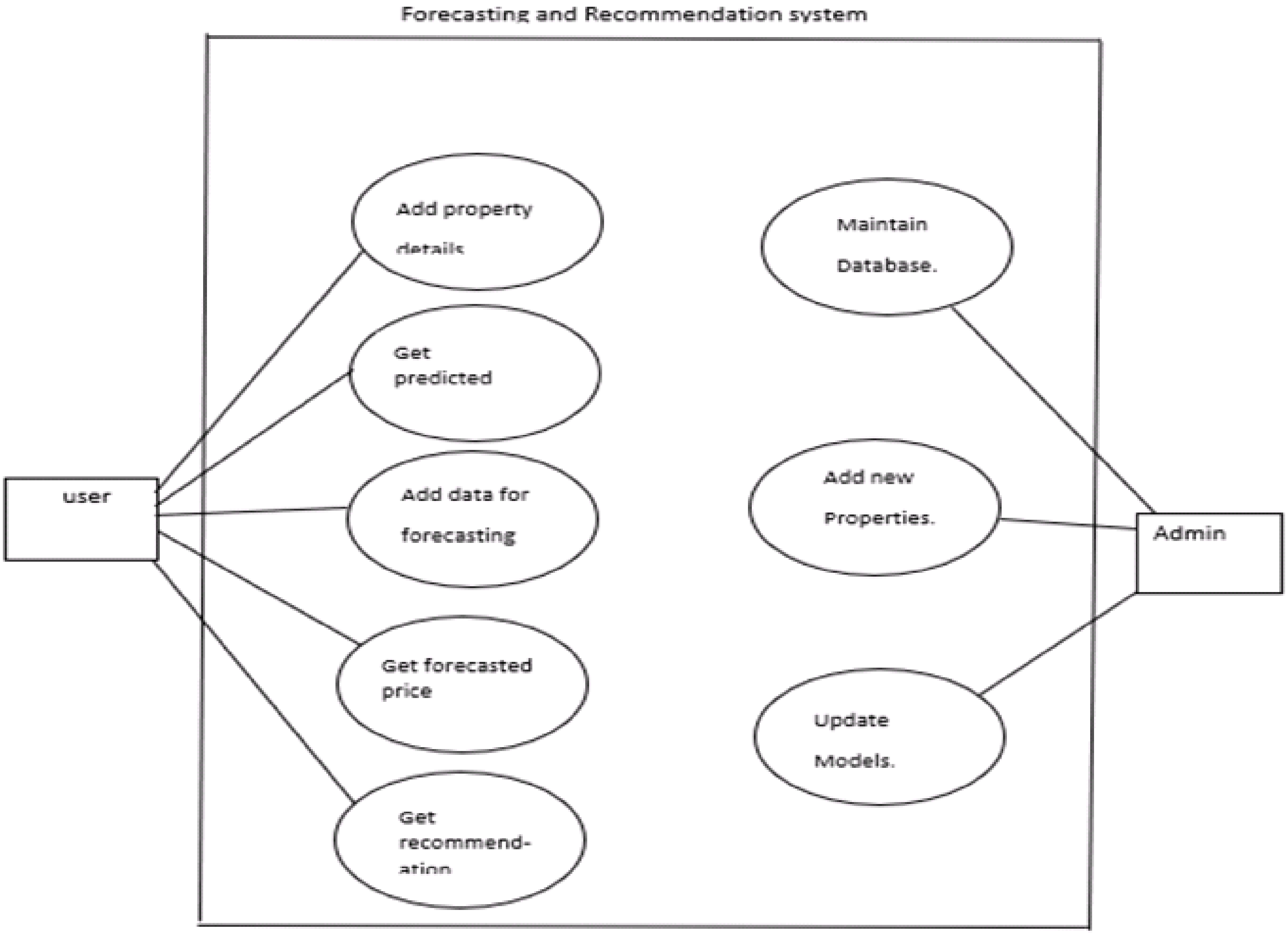


Fig -2: User activity diagram

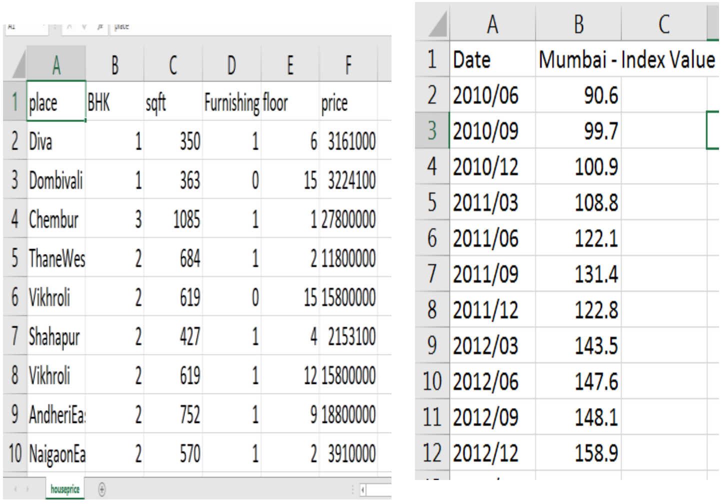
5. DESIGN AND IMPLEMENTATION

5.1 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.

5.2 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.



	A	B	C	D	E	F
1	place	BHK	sqft	Furnishing	floor	price
2	Diva	1	350	1	6	3161000
3	Dombivali	1	363	0	15	3224100
4	Chembur	3	1085	1	1	27800000
5	ThaneWes	2	684	1	2	11800000
6	Vikhroli	2	619	0	15	15800000
7	Shahapur	2	427	1	4	2153100
8	Vikhroli	2	619	1	12	15800000
9	AndheriEa:	2	752	1	9	18800000
10	NaigaonEa	2	570	1	2	3910000

	A	B	C
1	Date	Mumbai - Index Value	
2	2010/06	90.6	
3	2010/09	99.7	
4	2010/12	100.9	
5	2011/03	108.8	
6	2011/06	122.1	
7	2011/09	131.4	
8	2011/12	122.8	
9	2012/03	143.5	
10	2012/06	147.6	
11	2012/09	148.1	
12	2012/12	158.9	

Fig -3: Datasets

5.3 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.

5.4 Methodology:

5.4.1 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%.
- Formula:

$$Yi=\beta0+\beta1Xi1+\beta2Xi2+...+\beta pxi+ \epsilon$$

Yi=dependent variable

Xi=independent variables

$\beta0$= y-intercept (constant term)

ϵ=the model's error

INDEPENDENT VARIABLES	DEPENDENT VARIABLE
LOCATION (STRING)	PRICE (rupees)
BHK (INT)	
FURNISHING (0/1)	
SQ.FT (INT)	
OLD (INT)	
FLOOR (INT)	

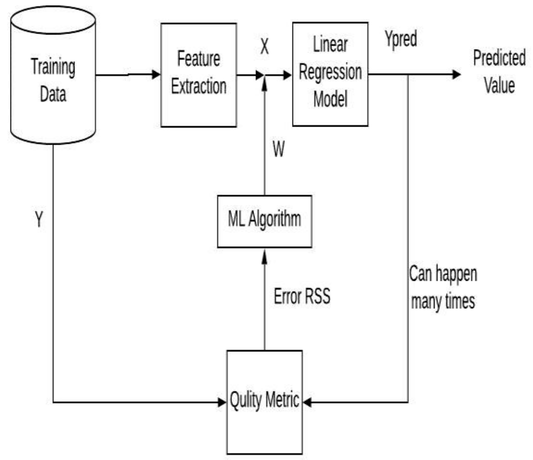


Fig -4: Prediction Model

5.4.2 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.

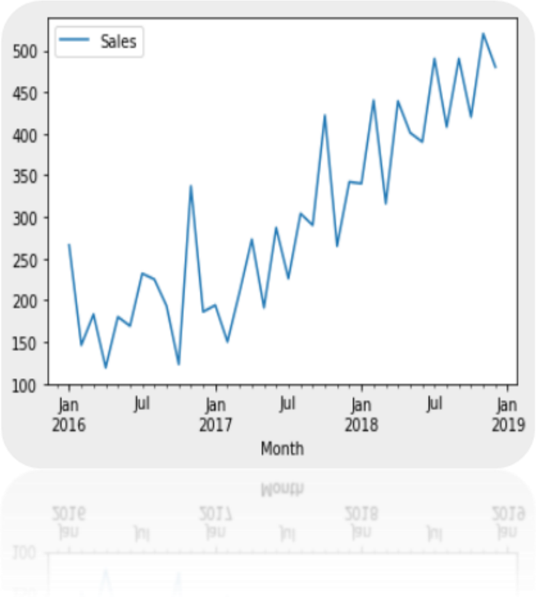


Fig-5:Non Stationary

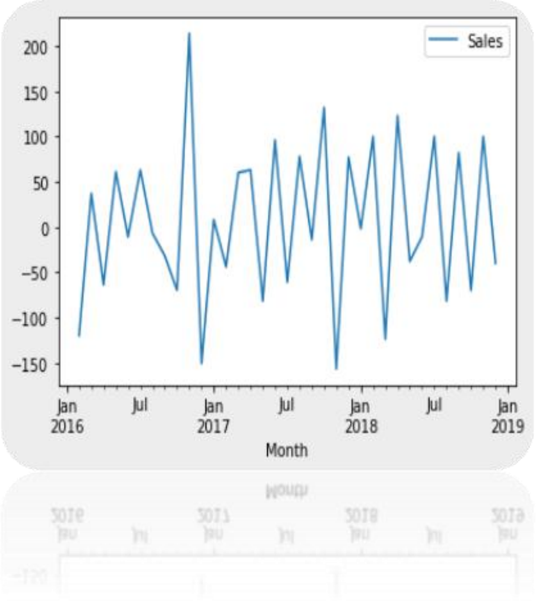


Fig -6:Stationary

- 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.

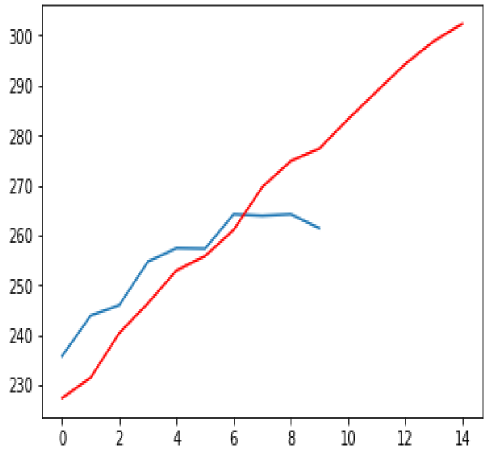
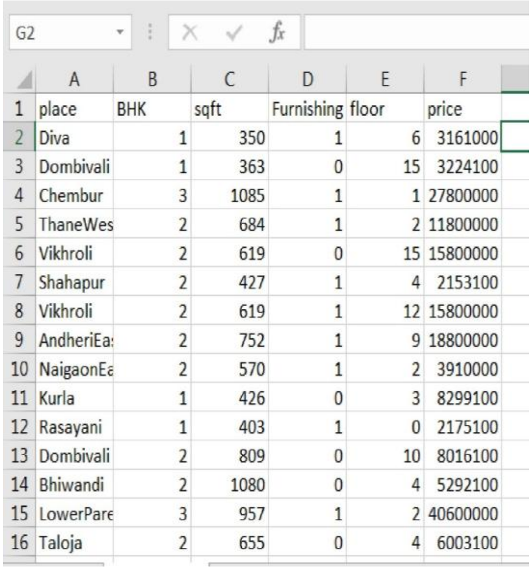


Chart -1: Accuracy of Arima Model

5.4.2 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.



	A	B	C	D	E	F
1	place	BHK	sqft	Furnishing	floor	price
2	Diva	1	350	1	6	3161000
3	Dombivali	1	363	0	15	3224100
4	Chembur	3	1085	1	1	27800000
5	ThaneWes	2	684	1	2	11800000
6	Vikhroli	2	619	0	15	15800000
7	Shahapur	2	427	1	4	2153100
8	Vikhroli	2	619	1	12	15800000
9	AndheriEa:	2	752	1	9	18800000
10	NaigaonEa	2	570	1	2	3910000
11	Kurla	1	426	0	3	8299100
12	Rasayani	1	403	1	0	2175100
13	Dombivali	2	809	0	10	8016100
14	Bhiwandi	2	1080	0	4	5292100
15	LowerPare	3	957	1	2	40600000
16	Taloja	2	655	0	4	6003100

Fig -7: Dataset for Prediction

- 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.

5.5 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.

5.6 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.

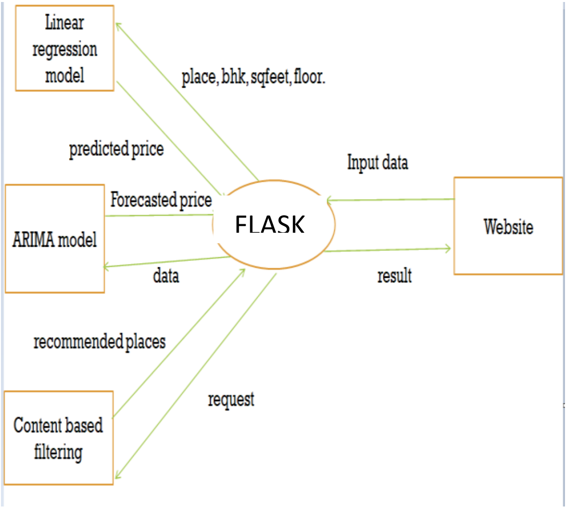


Fig -8: Working of System

- 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.

6. TECHNOLOGY USED

Documentation Tools

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

Language Used

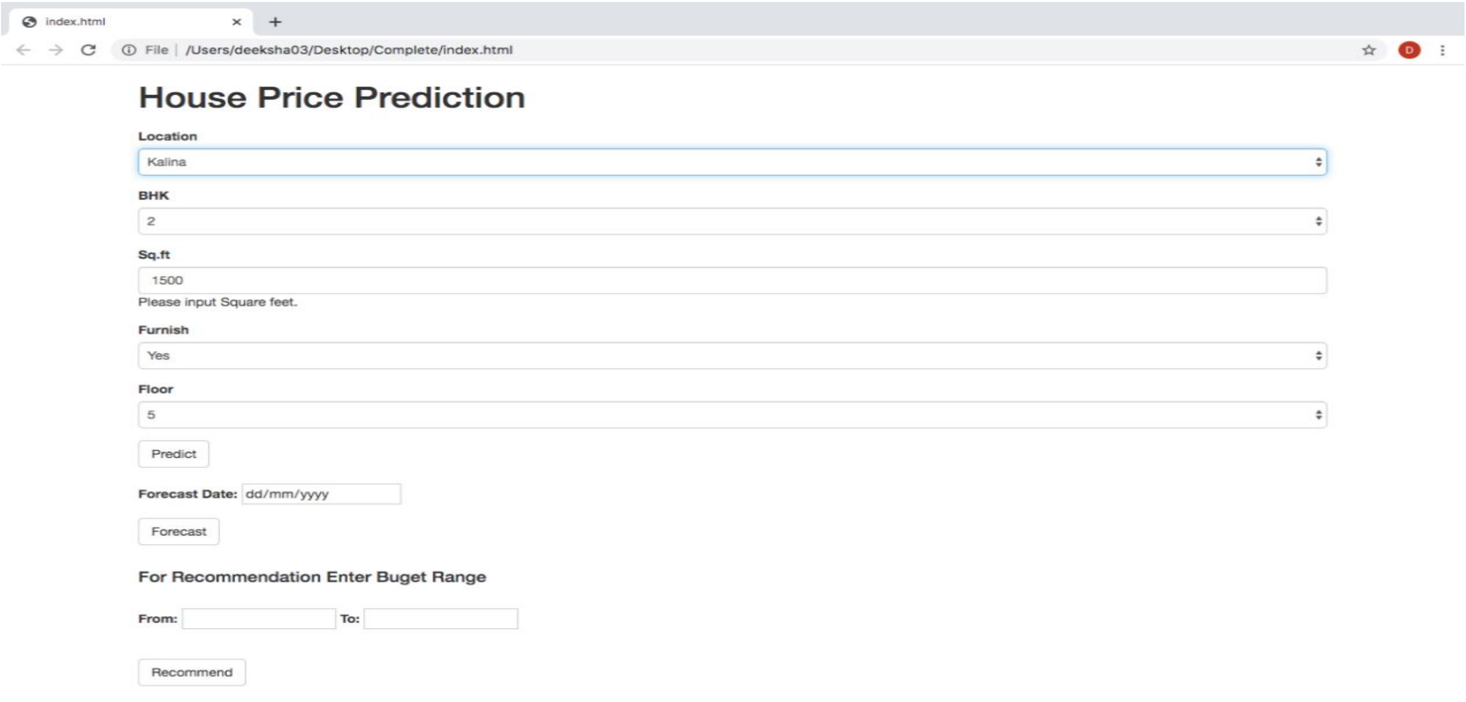
The "**HOUSE PRICE FORECASTING AND RECOMMENDATION SYSTEM**" will be used to for predicting the house price, forecasting that price and also to get best recommendation according to users requirement. This application can be run using website.

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.

7.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.



House Price Prediction

Location: Kalina

BHK: 2

Sq.ft: 1500

Please input Square feet.

Furnish: Yes

Floor: 5

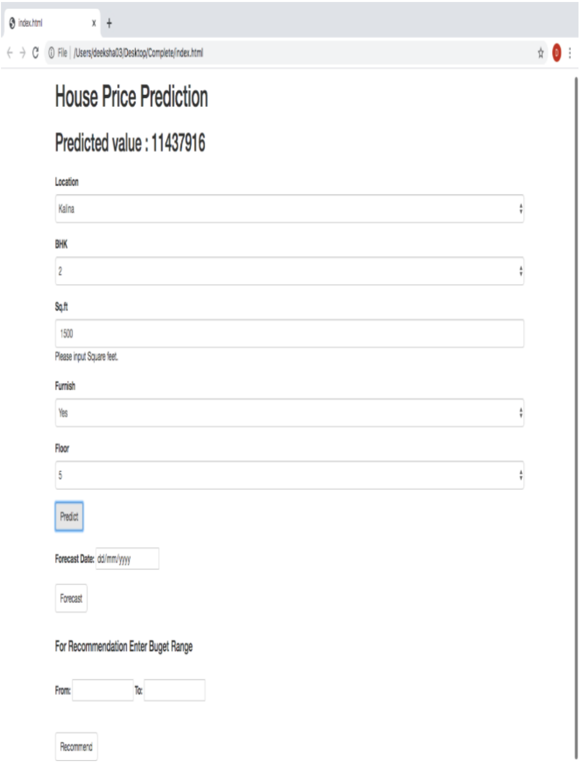
Forecast Date: dd/mm/yyyy

For Recommendation Enter Buget Range

From: To:

Fig -9: Front Webpage

After click on predict button website will show an estimated price of house according to input given by the user.



House Price Prediction

Predicted value : 11437916

Location: Kalina

BHK: 2

Sq.ft: 1500

Please input Square feet.

Furnish: Yes

Floor: 5

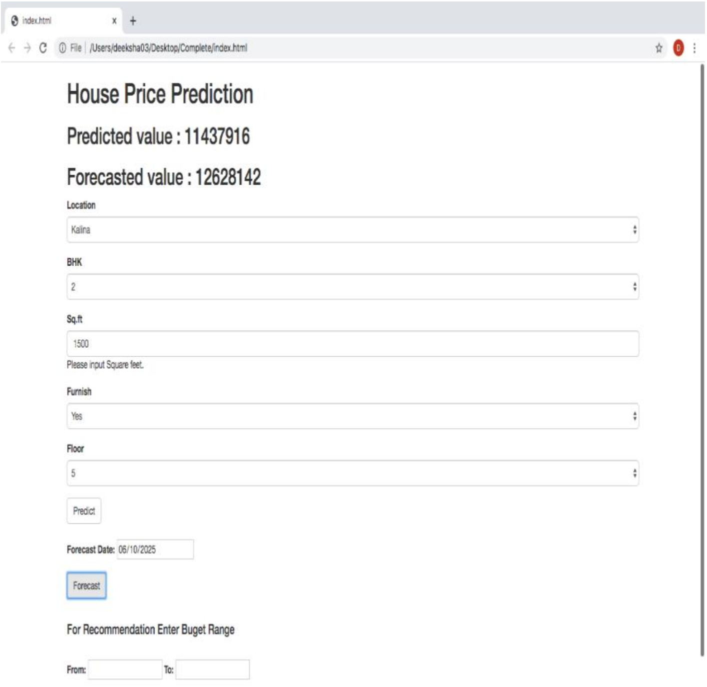
Forecast Date: dd/mm/yyyy

For Recommendation Enter Buget Range

From: To:

Fig -10: Result after Prediction

For forecasting we have enter date at which we want to know the price of house (after few year), after click on forecast button website will show an estimated price of house at that future date.



House Price Prediction

Predicted value : 11437916

Forecasted value : 12628142

Location: Kalina

BHK: 2

Sq.ft: 1500

Please input Square feet.

Furnish: Yes

Floor: 5

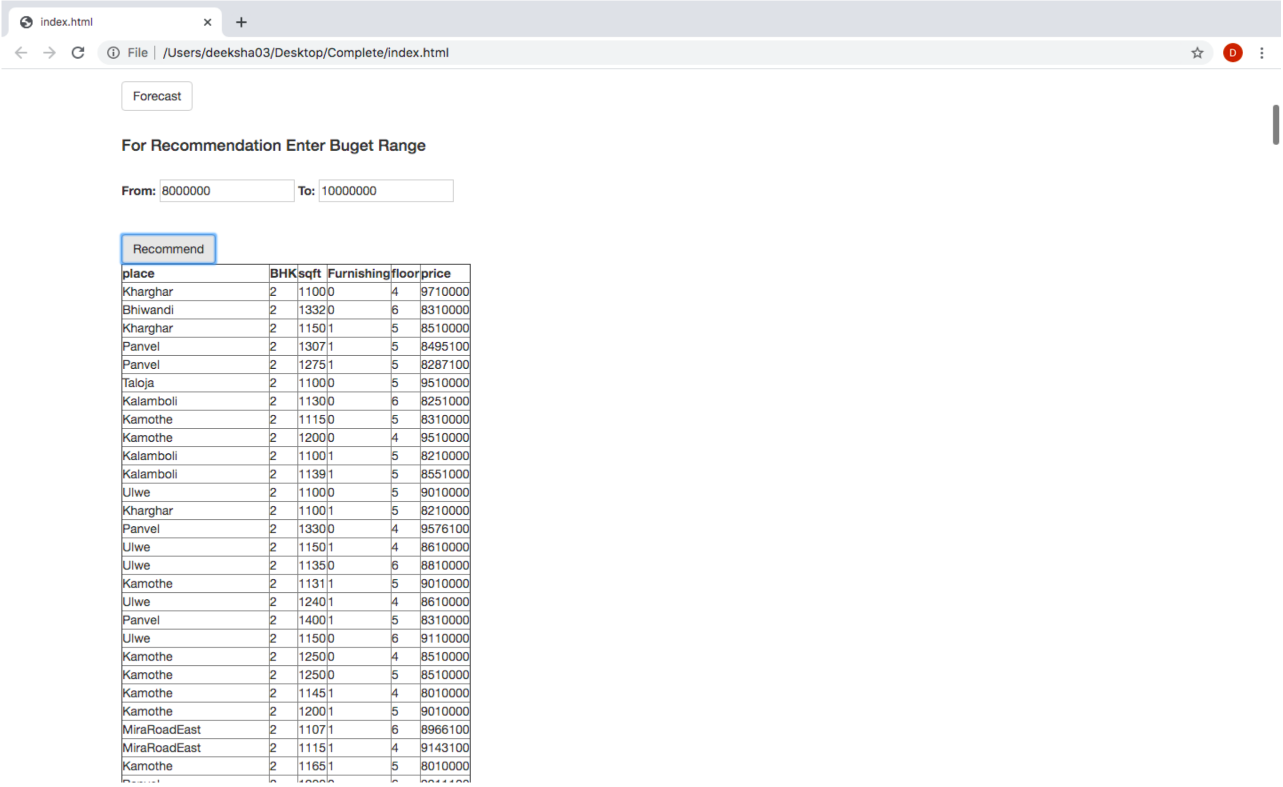
Forecast Date: 06/10/2025

For Recommendation Enter Buget Range

From: To:

Fig -11: Result after forecasting

If user wants recommendation for best house location according to his choice then he have to enter its budget from low to high after click on recommend button website will show the best house location according to input given by the user.



For Recommendation Enter Buget Range

From: 8000000 To: 10000000

place	BHK	sqft	Furnishing	floor	price
Kharghar	2	1100	0	4	9710000
Bhiwandi	2	1332	0	6	8310000
Kharghar	2	1150	1	5	8510000
Panvel	2	1307	1	5	8495100
Panvel	2	1275	1	5	8287100
Taloja	2	1100	0	5	9510000
Kalamboli	2	1130	0	6	8251000
Kamothe	2	1115	0	5	8310000
Kamothe	2	1200	0	4	9510000
Kalamboli	2	1100	1	5	8210000
Kalamboli	2	1139	1	5	8551000
Ulwe	2	1100	0	5	9010000
Kharghar	2	1100	1	5	8210000
Panvel	2	1330	0	4	9576100
Ulwe	2	1150	1	4	8610000
Ulwe	2	1135	0	6	8810000
Kamothe	2	1131	1	5	9010000
Ulwe	2	1240	1	4	8610000
Panvel	2	1400	1	5	8310000
Ulwe	2	1150	0	6	9110000
Kamothe	2	1250	0	4	8510000
Kamothe	2	1250	0	5	8510000
Kamothe	2	1145	1	4	8010000
Kamothe	2	1200	1	5	9010000
MiraRoadEast	2	1107	1	6	8966100
MiraRoadEast	2	1115	1	4	9143100
Kamothe	2	1165	1	5	8010000

Fig -12: Result after Recommendation

8. 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.

9. REFERENCES

[1] Real Estate Price Prediction with Regression and Classification, CS 229 Autumn 2016

[2] Gongzhu Hu, Jinping Wang, and Wenying Feng Multivariate Regression Modelling for Home Value Estimates with Evaluation using Maximum Information Coefficient

[3] Byeonghwa Park, Jae Kwon Bae (2015). Using machine learning algorithms for housing price prediction, Volume 42, Pages 2928-2934

[4] https://www.coursera.org/specializations/recommender-systems

[5] https://www.analyticsvidhya.com/blog/2015/08/beginners-guide-learn-content-based-recommender-systems/

[6] https://towardsdatascience.com/how-to-build-from-scratch-a-content-based-movie-recommender-with-natural-language-processing-25ad400eb243

[7] https://www.makaan.com/

[8] https://tradingeconomics.com/