International Journal of Future Generation Communication and Networking
Vol.14, No. 1, (2021), pp. 2428–2444

Property Price Prediction Application developed using Machine Learning Algorithm

Kulaye Shreyal Ashok[1*], Shaikh Mohammad Bilal Naseem[-2]
[1]Department of Computer Science
Somaiya Vidyavihar University Mumbai, India
[2]Assistant Professor, Dept of Computer Science/IT
Somaiya Vidyavihar University Mumbai, India
[1]shreyalkulaye17@gmail.com, [2]mohammadbilal@somaiya.edu

Abstract

Machine Learning is a Widely Growing Technology which has several highly fascinating practical applications that drive the kind of tasks and real business results. In the present paper we will discuss about the prediction of future Real Estate Housing Prices that is generated and developed by machine learning algorithm and how to connect machine learning model to Application Programming Interface. Property prices are increasing vigorously every year, so there is a need and high scope for a system to predict real estate prices in future. User interface will help to connect to a machine learning model which gets easier for lay public and connection of machine learning model to application programming interface will be useful for making interactive programming interfaces with higher functionality.

Keywords: *Machine Learning, Linear Regression, Property Price Prediction, Real Estate, Lasso, Decision Tree, Supervised Learning, House Price Prediction Application.*

1. Introduction

We human being learns every minute something new based on our past experiences, mistakes and from current scenarios in our lives and make some conclusions and predictions on the basis of that experiences. That's what learning actually mean. Similarly, In the machine learning past data and current scenarios plays a vital role same as human beings uses past experiences for acquiring and expertise in current situations and make solutions. With the inspiration of the real world problems and decision making we demonstrate a typical machine learning model which can give solutions on past experiences and based on current situations which will predict the solutions to the problem. It makes an easier for rational decision making[1].

This study utilizes machine learning algorithms as a method of research which develops prediction of property prices. For the selection of prediction methodology we explore and compare various prediction methods and models. We have implemented Supervised Learning strategy to train and test the data model[2]. We have compared prediction results with linear regression, lasso as well as decision tree models. We utilized Linear Regression Machine Learning algorithm as our model of prediction because it's best fitted, probabilistic and highly adaptable methodology on model selection for prediction. Property price prediction model is recommended in a part to support a property purchaser, house seeker, real estate agent or a house vendor for information based on valuation of a particular house/flat in particular area.

Every day real estate prices are getting high and high. As real estate is a very important field in investment, we have to take proper and rational decisions[4]. We have

ISSN: 2233-7857IJFGCN

developed proper mechanism which can give us predictions. This type of prediction mechanism and proper predictions can be achieved by machine learning algorithms[5]. That's the scope of machine learning in the present world and the coming future which will help to take intelligent decisions. As general people who doesn't know much about programming, we have developed a machine learning based application via python flask server[10]. Python flask server is working as a backend and serving all the solutions for the operations invoked. Application programming interface is working as simple user interface which will serve results and solutions for the functions. So it get easier for understanding solution for the real life problems.

2. Problem Statement

A. Statement of Problem

To analyze the given data and compare the algorithms and then predict the house prices on the basis of best fit machine learning model and developing application programming interface for the same.

There are so many types of model which we can go for. For property price prediction regression models suits the best. Basically, three types of regression models can fit best for this property price prediction model, those are linear regression, decision tree and lasso regression. All the three again vary in performance according to model and data provided. So we have chosen one with high accuracy rate from them again and then we have carry forwarded that for best prediction results[11].

B. Existing Systems

The existing system involves calculation of present prices of properties without the necessary price prediction about the future increasing market prices and trends. so as to overcome this fault, there is need of an updated and automated system which can predict future real estate prices. Also there are so many models based on prediction in machine learning and data mining which gives proper predictions, but there is need of user interface for general people who doesn't know machine learning or any other programming language and it's functionality[7]. As everything is shifting from manual to highly automated digital systems like simple application based and user friendly, There is a high need of some user interface or application interface which will serve basic needs and prediction which will be user friendly and simply understandable by people who doesn't know much about programming.

ISSN: 2233-7857IJFGCN

Literature Survey

Table No. 1

Sr. No.	*Paper Title*	*Author*	*Year*	*Advantages*	*Limitations*
1.	***Prediction of Residential Property Prices – A State of the Art***	*Rohan Bafna1, Anirudh Dhole2, Ankit Jagtap3, Asif Kazi4, Arbaz Kazi*	*2018*	*Importance of developing middle man trusted application [7]*	*Strategies are mentioned but lacking in giving idea of implementation.*
2.	***House Price Prediction Using Machine Learning***	*G. Naga Satish, Ch.V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu*	*2019*	*Gradient Boosting algorithm strategy [5]*	*Proper explanation of example was needed to give justification of chosen algorithm.*
3.	***House Price Forecasting using Data Mining***	*Nihar Bhagat, Ankit Mohokar, Shreyash Mane*	*2016*	*Data mining techniques and idea of GUI [6]*	*No accuracy measure is provided of the algorithm used.*
4.	***Machine Learning Housing Price Prediction in Petaling Jaya, Selangor, Malaysia***	*Thuraiya Mohd, Suraya Masrom, Noraini Johari*	*2019*	*Comparison between different regression algorithms.[8]*	*Not giving implementation of best predicting model.*
5.	***The Application Of Data Mining Technology In Real Estate Market Prediction***	*Xian Guang LI, Qi Ming LI*	*2006*	*Stated data mining technologies and it's applications in real estate industry.[13]*	*Different technologies are mentioned but lacking in idea about choosing proper algorithm and implementation.*
6.	***Real Estate Price Prediction Using Machine Learning***	*Aswin Sivam Ravikumar*	*2018*	*Explained proper architecture and mentioned gradient boosting algorithm.[14]*	*Prediction result is not provided after choosing best fit algorithm.*

ISSN: 2233-7857IJFGCN

7.	***House Price Forecasting Using Data Mining***	*Nihar Bhagat, Ankit Mohokar, Shreyash Mane*	*2016*	*Given proper flow of working of the system.[15]*	*Haven't performed comparison between different algorithms*

Wei Luo in **"Guidelines for Developing and Reporting Machine Learning Predictive Models In Biomedical Research: A Multidisciplinary View"** mentioned steps for implementing a machine learning model[16] which gives proper idea for building a machine learning model. Also Djamael Djenouri in **"Machine Learning For Smart Building Applications: Review And Taxonomy"** explained framework for building smart machine learning applications[17]. Xian Guang LI, Qi Ming LI in **"The Application Of Data Mining Technology In Real Estate Market Prediction"** stated various data mining technologies which can help in building machine learning model and also stated applications of data mining in real estate industry[13]. Nihar Bhagat, Ankit Mohokar, Shreyash Mane in **"House Price Forecasting Using Data Mining"** explained working of the system with the flow diagram which is giving idea of implementation of steps for the model building[15]**.** Aswin Sivam Ravikumar in **"Real Estate Price Prediction Using Machine Learning"** utilized gradient boosting algorithm for the model selection and explained proper architecture of the machine learning model[14].

As shown in Table 1, Rohan Bafna, Anirudh Dhole2, Ankit Jagtap3, Asif Kazi4, Arbaz Kazi in **"Prediction of Residential Property Prices – A State of the Art"** mentioned Literature Review which gives idea about upcoming strategies also explained need of trustful interface for prediction criteria[7]. G. Naga Satish, Ch.V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu in "**House Price Prediction Using Machine Learning"** stated implementation in detailed manner with graphs also explained importance of Gradient Boosting Algorithm and it is giving 91.14233 accuracy rate[5]. Nihar Bhagat, Ankit Mohokar, Shreyash Mane in **"House Price Forecasting using Data Mining"** discussed about linear regression it's importance and gives idea about the interface like GUI[6]. Thuraiya Mohd, Suraya Masrom, Noraini Johari in **"Machine Learning Housing Price Prediction in Petaling Jaya, Selangor, Malaysia"** provided accuracy rate of various machine learning algorithms. There comes two most important things which decides performance of particular prediction model first is accuracy of prediction algorithm and second is averaging errors of fitness[8].

3. Proposed Model

For building a machine learning model we first have to appropriately define the problem. The very first and most critical thing to do is to identify what will be the inputs and what is the expected output. In this property price prediction model our main objective is to predict future property prices. Target features are based on our dataset like total square fit, BHK(Bedroom, Hall, Kitchen), location, price, number of bathrooms, number of balcony, etc. We have wisely chosen inputs from the dataset. For selection of input we will mainly focus on features which are dependent on model to predict prices. We have analyzed that our available data is sufficient informative to learn the relationship between the inputs and the output. Now, let us look at steps for building machine learning model for property price prediction.

ISSN: 2233-7857IJFGCN

Steps for building machine learning model

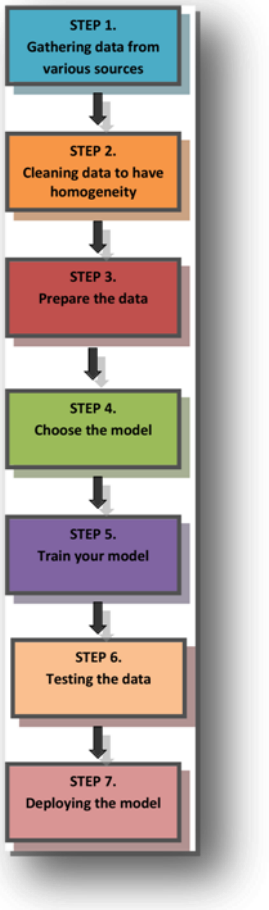


Figure No. 1

Step 1. Data Collection

Collecting data is the first step towards the real development of our machine learning model. This is the step which actually cascades how good our model will be. The more and better data we get, the better our models prediction result will be. Data plays vital role in performance of the model and it's result.

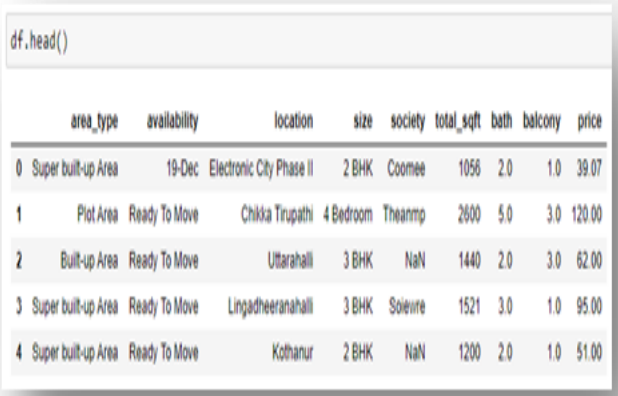
area_typ	availabili	location	size	society	total_sq	bath	balcony	price
Super bu	19-Dec	Electron	2 BHK	Coomee	1056	2	1	39.07
Plot Are	Ready T	Chikka T	4 Bedro	Theanm	2600	5	3	120
Built-up	Ready T	Uttaraha	3 BHK		1440	2	3	62
Super bu	Ready T	Lingadhe	3 BHK	Soiewre	1521	3	1	95
Super bu	Ready T	Kothanu	2 BHK		1200	2	1	51
Super bu	Ready T	Whitefiel	2 BHK	DuenaT.	1170	2	1	38
Super bu	18-May	Old Airp	4 BHK	Jaades	2732	4		204
Super bu	Ready T	Rajaji Na	4 BHK	Brway G	3300	4		600
Super bu	Ready T	Maratha	3 BHK		1310	3	1	63.25
Plot Are	Ready T	Gandhi E	6 Bedroom		1020	6		370
Super bu	18-Feb	Whitefiel	3 BHK		1800	2	2	70
Plot Are	Ready T	Whitefiel	4 Bedro	Prrry M	2785	5	3	295
Super bu	Ready T	7th Phas	2 BHK	Shncyes	1000	2	1	38
Built-up	Ready T	Gottiger	2 BHK		1100	2	2	40
Plot Are	Ready T	Sarjapur	3 Bedro	Skityer	2250	3	2	148
Super bu	Ready T	Mysore	2 BHK	PrntaEn	1175	2	2	73.5
Super bu	Ready T	Bisuvan	3 BHK	Prityel	1180	3	2	48
Super bu	Ready T	Raja Raj	3 BHK	GrrvaGr	1540	3	3	60
Super bu	Ready T	Ramakri	3 BHK	PeBayle	2770	4	2	290

Figure No. 2

ISSN: 2233-7857IJFGCN

The previous table(fig. 2) corresponds to famous Bangalore house price dataset frequently used to develop machine learning model[12]. we have developed our property price prediction model using this dataset.

Next we have explored the dataset. After gathering the data exploration of the data is important step to perform. By exploring the dataset we get the idea about structure of the data. We analyze the volume of the by getting the values of number of rows and columns. We have to perform various operations on data, so it is required to get the data types of the columns in the data. By describing our data we got the values like count, mean, standard deviation, etc. By performing all this steps we get overall view of the data and it is helpful for the data cleaning step.

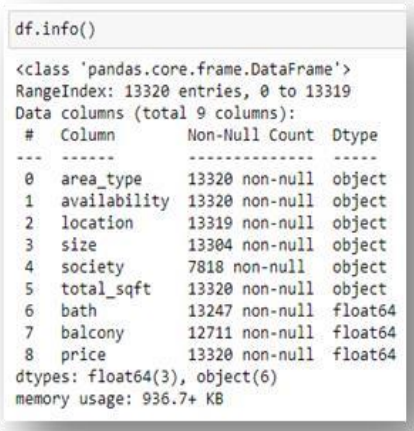


```
df.head()
```

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00

Figure No. 3

As shown in fig. 3, We have shown structure of the data by applying head function.



```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13320 entries, 0 to 13319
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   area_type     13320 non-null  object
 1   availability  13320 non-null  object
 2   location      13319 non-null  object
 3   size          13304 non-null  object
 4   society       7818 non-null   object
 5   total_sqft    13320 non-null  object
 6   bath          13247 non-null  float64
 7   balcony       12711 non-null  float64
 8   price         13320 non-null  float64
dtypes: float64(3), object(6)
memory usage: 936.7+ KB
```

Figure No. 4

For getting data types of the columns in the dataset we have performed info operation on the dataset which gives detailed information about the columns along with data type and count as shown in fig. 4.

ISSN: 2233-7857IJFGCN

```
df.describe()
```

	bath	balcony	price
count	13247.000000	12711.000000	13320.000000
mean	2.692610	1.584376	112.565627
std	1.341458	0.817263	148.971674
min	1.000000	0.000000	8.000000
25%	2.000000	1.000000	50.000000
50%	2.000000	2.000000	72.000000
75%	3.000000	2.000000	120.000000
max	40.000000	3.000000	3600.000000

```
df.shape
```

```
(13320, 9)
```

Figure No. 5

As shown in fig. 5, We got values of count, mean, standard deviation by applying describe operation which will be helpful for further evaluation.

Step 2. Data Cleaning

After gathering dataset the next important step is to clean the data. Data cleaning is the process of detecting and correcting inaccurate records. For cleaning the data first we have explored the dataset. We identified which features are important for prediction and we have kept only those for prediction criteria. In our dataset we have dropped area type, society, balcony and availability columns. we have identified null values and we have dropped them as shown below in fig. 6.

```
df1 = df.drop(['area_type','society','balcony','availability'],axis='columns')
```

```
df1.head()
```

	location	size	total_sqft	bath	price
0	Electronic City Phase II	2 BHK	1056	2.0	39.07
1	Chikka Tirupathi	4 Bedroom	2600	5.0	120.00
2	Uttarahalli	3 BHK	1440	2.0	62.00
3	Lingadheeranahalli	3 BHK	1521	3.0	95.00
4	Kothanur	2 BHK	1200	2.0	51.00

Figure No. 6

Step 3. Prepare The Data

Now after cleaning the data it is very important to keep the data in homogeneous format. So by applying various functions we have converted the data into homogeneous format. After preparing the data in homogeneous format next step is to detect the outliers and remove them.

ISSN: 2233-7857IJFGCN

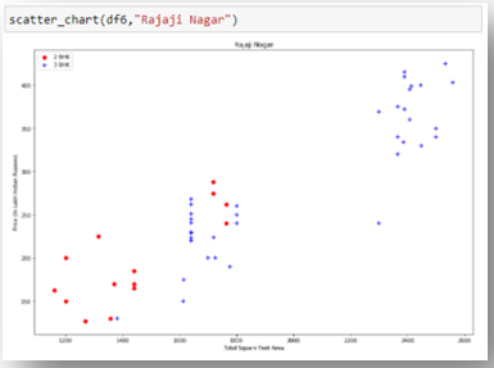


Figure No. 7

```
def remove_bhk_outliers(df):
    exclude_indices = np.array([])
    for location, location_df in df.groupby('location'):
        bhk_stats = {}
        for bhk, bhk_df in location_df.groupby('bhk'):
            bhk_stats[bhk] = {
                'mean': np.mean(bhk_df.price_per_sqft),
                'std': np.std(bhk_df.price_per_sqft),
                'count': bhk_df.shape[0]
            }
        for bhk, bhk_df in location_df.groupby('bhk'):
            stats = bhk_stats.get(bhk-1)
            if stats and stats['count']>5:
                exclude_indices = np.append(exclude_indices, bhk_df[bhk_df.price_per_sqft<(stats['mean'])].index.values)
    return df.drop(exclude_indices,axis='index')
```

```
df7 = remove_bhk_outliers(df6)
df7.shape
```

```
(7329, 7)
```

Figure No. 8

As shown in fig. 8 we have developed a function which can remove outliers from the data. By applying above function we have removed outliers from the data which can create anomalies in prediction of prices.

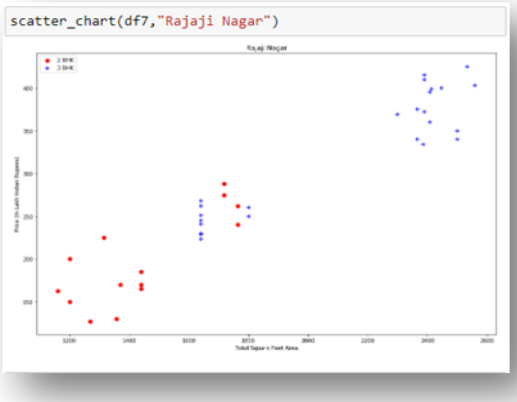


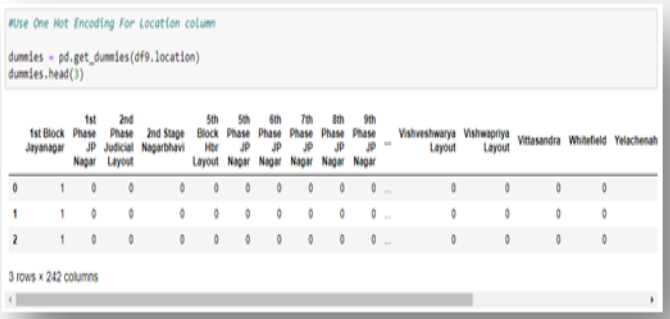
Figure No. 9

We have analyzed the outliers and removed them by applying functions as shown in fig. 8. In fig. 7 we have detected the outlier and in fig. 9 by applying function we have removed the outliers.

ISSN: 2233-7857IJFGCN

Step 4. Choose The Model

After cleaning and preparing the data the next step is to convert categorical data into numeric data. For that we have used One Hot Encoding technique as shown in the fig. 10 below.



```
#Use One Hot Encoding For Location column

dummies = pd.get_dummies(df9.location)
dummies.head(3)
```

	1st Block Jayanagar	1st Phase JP Nagar	2nd Phase Judicial Layout	2nd Stage Nagarbhavi	5th Block Hbr Layout	5th Phase JP Nagar	6th Phase JP Nagar	7th Phase JP Nagar	8th Phase JP Nagar	9th Phase JP Nagar	...	Vishveshwarya Layout	Vishwapriya Layout	Vittasandra	Whitefield	Yelachenah
0	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
1	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
2	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	

3 rows × 242 columns

Figure No. 10

After converting the data into numerical format next step is to choose the model. For choosing the model we have implemented GridSearchCV[11] algorithm.

```
#Using GridSearchCV Model

from sklearn.model_selection import GridSearchCV

from sklearn.linear_model import Lasso
from sklearn.tree import DecisionTreeRegressor

def find_best_model_using_gridsearchcv(X,Y):
    algos = {
        'linear_regression' : {
            'model': LinearRegression(),
            'params': {
                'normalize': [True, False]
            }
        },
        'lasso': {
            'model': Lasso(),'params': {
                'alpha': [1,2],
                'selection': ['random', 'cyclic']
            }
        },
        'decision_tree': {
            'model': DecisionTreeRegressor(),
            'params': {
                'criterion' : ['mse','friedman_mse'],
                'splitter': ['best','random']
            }
        } }
```

Figure No. 11

For prediction machine learning models regression algorithms works the best and gives proper estimation. Among various regression models three regression algorithms give high accuracy for prediction models. Those three algorithms of regression are Linear Regression, Lasso Regression and Decision Tree Regression. We have compared accuracy scores of these three algorithms and one with highest score we have chosen for our prediction results.

ISSN: 2233-7857IJFGCN

```
    scores = []
    cv = ShuffleSplit(n_splits=5, test_size=0.2, random_state=0)
    for algo_name, config in algos.items():
        gs =  GridSearchCV(config['model'], config['params'], cv=cv, return_train_score=False)
        gs.fit(X,Y)
        scores.append({
            'model': algo_name,
            'best_score': gs.best_score_,
            'best_params': gs.best_params_
        })

    return pd.DataFrame(scores,columns=['model','best_score','best_params'])

find_best_model_using_gridsearchcv(X,Y)
```

	model	best_score	best_params
0	linear_regression	0.818354	{'normalize': False}
1	lasso	0.687429	{'alpha': 1, 'selection': 'cyclic'}
2	decision_tree	0.725400	{'criterion': 'mse', 'splitter': 'best'}

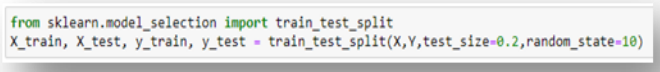
```
#Based on above results we can say that LinearRegression gives the best score. Hence we will use that.
```

Figure No. 12

Results of GridSearchCV algorithm are as shown in figure number 12. By implementing GridSearchCV we have calculated accuracy score for linear regression, lasso regression and decision tree regression. Score for linear regression is nearby 82%. Score for lasso regression is approximately equal to 69% and score for decision tree regression is 72%. There comes two most important things which decides performance of particular prediction model first is accuracy of prediction algorithm and second is averaging errors of fitness. According to this result of GridSearchCV, linear regression is fitting best for this model. So we have utilized linear regression model because it is best fitted and has high accuracy rate.

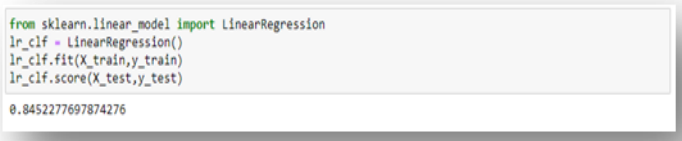
Step 5. Train Your Model

For training the data we have worked with linear regression model. Linear regression is relationship between a scalar response and one or more explanatory variables. For training the model we have assigned x and y variables. In x variable we have stored values of total square feet, BHK, bathroom, location and in y variable we have stored property prices of various locations. Then we have trained the data using scikit learn module [9] as shown in figure no. 13.



```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,Y,test_size=0.2,random_state=10)
```

Figure No. 13



```
from sklearn.linear_model import LinearRegression
lr_clf = LinearRegression()
lr_clf.fit(X_train,y_train)
lr_clf.score(X_test,y_test)
```

```
0.8452277697874276
```

Figure No. 14

We have performed linear classification training on the data which is of the type linear regression[18] and we have again checked the score. Here again the score for linear regression is 84% as shown in fig. 14.

ISSN: 2233-7857IJFGCN

```
#Using K Fold cross validation to measure accuracy of our LinearRegression model

from sklearn.model_selection import ShuffleSplit
from sklearn.model_selection import cross_val_score

cv = ShuffleSplit(n_splits=5, test_size=0.2, random_state=0)

cross_val_score(LinearRegression(), X, Y, cv=cv)
```

```
array([0.82430186, 0.77166234, 0.85089567, 0.80837764, 0.83653286])
```

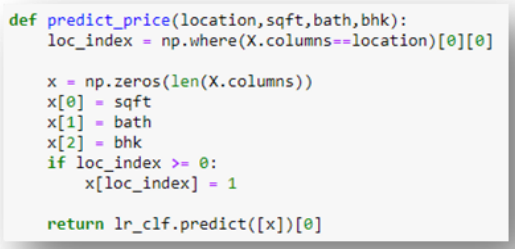
```
#We can see that in 5 iterations we get a score above 80% all the time. This is pretty good but I am testing few other algorithms
```

Figure No. 15

we have applied K Fold Cross validation technique to measure accuracy of our linear regression model[19]. We have applied shufflesplit operation to shuffle the data and performed cross value score operation on the data for random 5 iterations. According to fig. 15 we can see that we got the scorers for 5 iterations like 82.43%, 77.16%, 85.08%, 80.83% and 83.86%. We can state that we get the score almost above 80% all the time. This proves that we have chosen pretty good fitting algorithm for our prediction model which will give high accuracy for predicted prices.

Step 6. Testing The Data

For testing the data we have developed a function to predict price as shown below in fig. 16 and 17.

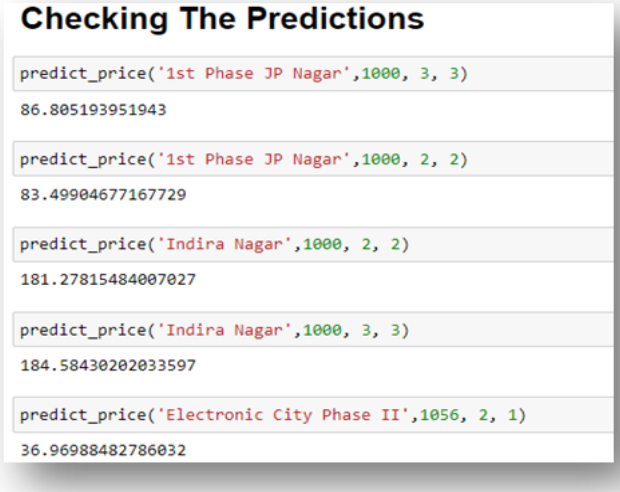


```
def predict_price(location,sqft,bath,bhk):
    loc_index = np.where(X.columns==location)[0][0]

    x = np.zeros(len(X.columns))
    x[0] = sqft
    x[1] = bath
    x[2] = bhk
    if loc_index >= 0:
        x[loc_index] = 1

    return lr_clf.predict([x])[0]
```

Figure No. 16



Checking The Predictions

```
predict_price('1st Phase JP Nagar',1000, 3, 3)
```

```
86.805193951943
```

```
predict_price('1st Phase JP Nagar',1000, 2, 2)
```

```
83.49904677167729
```

```
predict_price('Indira Nagar',1000, 2, 2)
```

```
181.27815484007027
```

```
predict_price('Indira Nagar',1000, 3, 3)
```

```
184.58430202033597
```

```
predict_price('Electronic City Phase II',1056, 2, 1)
```

```
36.96988482786032
```

Figure No. 17

ISSN: 2233-7857IJFGCN

Step 7. Deploying the model

For deploying the model which we have built, we have created pickle file of the model as shown in fig. 18 and 19 and connected to python flask server[10] as shown in fig.20 and 21. For accepting the locations we have created javascript Object notation(.json) file because python file accepts the data in the form of javascript object notation(.json) format as shown in fig. 18 and 19. For using our model in application, we have imported the pickle file in artifacts for utilizing functions created in our main model from pickle file as shown in fig. 21. After loading artifacts we have connected it to python flask server[10] as shown in fig. 20. Python flask server is working as backend and providing service for our application. Frontend is connected to application programming interface using HTML page[3] and designing is done using CSS. In the frontend of the application we have kept fields for input like total square feet, BHK, bathroom and location. Then there is estimate price button control. On click of button the function written in backend revokes the method of price prediction and according to that function, price is estimated. The model of webapplication is shown below in fig. 22 and implementation of python flask server is shown in fig. 20 and 21.

```
import pickle
with open('House_Price_Prediction.pickle','wb') as f:
    pickle.dump(lr_clf,f)

#Exporting location and column information to a file that will be useful later on in our prediction application
import json
columns = {
    'data_columns' : [col.lower() for col in X.columns]
}
with open("columns.json","w") as f:
    f.write(json.dumps(columns))
```

Figure No. 18

Figure No. 19

```
from flask import Flask, request, jsonify
import util

app = Flask(__name__)

@app.route('/get_location_names', methods=['GET'])
def get_location_names():
    response = jsonify({
        'locations': util.get_location_names()
    })
    response.headers.add('Access-Control-Allow-Origin', '*')

    return response

@app.route('/predict_home_price', methods=['GET', 'POST'])
def predict_home_price():
    total_sqft = float(request.form['total_sqft'])
    location = request.form['location']
    bhk = int(request.form['bhk'])
    bath = int(request.form['bath'])

    response = jsonify({
        'estimated_price': util.get_estimated_price(location,total_sqft,bhk,bath)
    })
    response.headers.add('Access-Control-Allow-Origin', '*')

    return response
```

Figure No. 20

ISSN: 2233-7857IJFGCN

```
def load_saved_artifacts():
    print("loading saved artifacts...start")
    global __data_columns
    global __locations

    with open("./artifacts/columns.json", "r") as f:
        __data_columns = json.load(f)['data_columns']
        __locations = __data_columns[3:]  # first 3 columns are sqft, bath, bhk

    global __model
    if __model is None:
        with open('./artifacts/House_Price_Prediction.pickle', 'rb') as f:
            __model = pickle.load(f)
    print("loading saved artifacts...done")

def get_location_names():
    return __locations

def get_data_columns():
    return __data_columns
```

Figure No. 21

Figure No. 22

5. Survey Analysis

We have conducted survey as a part of the project based on utilization of Application of property price prediction via Google Forms. The form was submitted to the interested property investors. Average audience targeted was 125 and age was in the range of 20 to 70. We have asked some questions regarding to the topic and the summary report of the responses is as shown in figures below in the form of pie-charts.
Responses for the following questions:

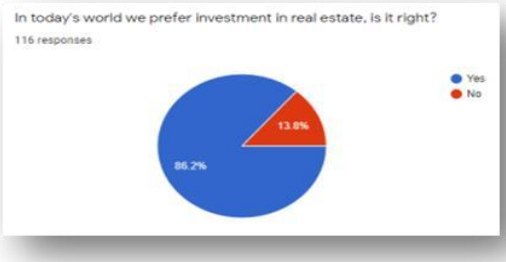


Figure No. 23

According to the fig. 23, 86.2% investors prefer investment in real estate.

ISSN: 2233-7857IJFGCN

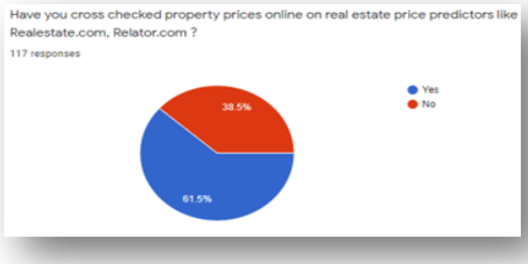


Figure No. 24

As shown in fig. 24, 61.5% investors cross check property prices online.

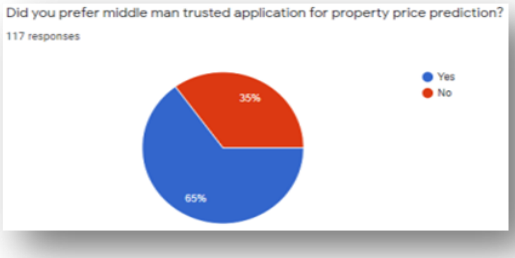


Figure No. 25

According to the response for question asked in fig. 25, 65% investors prefer middle man trusted application for property price prediction.

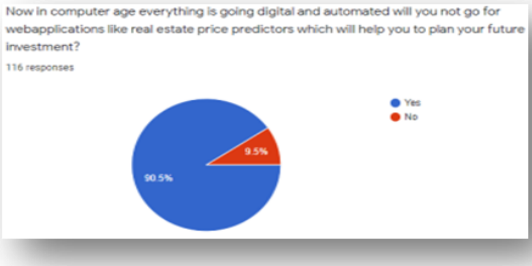


Figure No. 26

90.5% investors are with the logic planning before purchasing as shown in fig. 26.

Figure No. 27

97.5% agreed for first checking prices online and then go for purchase as shown in fig. 27.

ISSN: 2233-7857IJFGCN

According to the responses recorded from the survey it reveals that people prefer investing in real estate as it is safer and before purchasing property most of the people check the prices online because they tend to trust on 'middle man trusted application'. People supports first planning and then buying properties. Overall report concludes need and positive response for property price prediction application.

6. Conclusion

The main goal of this project is to provide application programming interface and predicting accurate future property prices. So the project is successfully implemented which is giving proper estimated property prices. Our research provides a way to compare various algorithms and gives proper idea to choose the best fitted algorithm. By the reference of our project anyone can successfully build machine learning model and can build user interface for society usage. This research will be helpful for society, government, data scientist, data analyst and who are interested in building machine learning models.

7. Acknowledgement

It is a matter of great honor to work on the research project on **"Property Price Prediction application developed using machine learning algorithm"**. The project received excellent guidance of project guide **Prof. Shaikh Mohammad Bilal N**. The project received their whole hearted assistance, inspiration, encouragement and valuable guidance in all phases.

8. References

[1] Simon Tavasoli,"The importance of machine learning", Simplilearn, Sep 25,2020 -https://www.simplilearn.com/importance-of-machine-learning-for-data-scientists-article

[2] Mark Ryan M. Talabis, D. Kaye, Information Security Analytics, "Supervised Learning", Sciencedirect, 2015 - https://www.sciencedirect.com/topics/computer-science/supervised-learning

[3] Sayak Paul, "Turning Machine learning models into API's", DataCamp, October 25 2018 - https://www.datacamp.com/community/tutorials/machine-learning-models-api-python

[4] The Oracles, "Real estate is still the best investment you can make today, millionaires say—here's why", Make It, Oct 1 2019 - https://www.cnbc.com/2019/10/01/real-estate-is-still-the-best-investment-you-can-make-today- millionaires-say.html

[5] G. Naga Satish, Ch. V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu, " House Price Prediction Using Machine Learning", IJITEE, 2019

[6] Nihar Bhagat, Ankit Mohokar, Shreyash Mane, " House Price Forecasting using Data Mining", IJCAONLINE, October 2016

[7] Rohan Bafna1, Anirudh Dhole2, Ankit Jagtap3, Asif Kazi4,Arbaz Kazi, " Prediction of Residential Property Prices – A State of the Art", ResearchGate, March 2018

[8] Thuraiya Mohd, Suraya Masrom, Noraini Johari, " Machine Learning Housing Price Prediction in Petaling Jaya, Selangor, Malaysia", IJRTE, September 2019

ISSN: 2233-7857IJFGCN

[9] Scikitlearn module, "Scikit Learn Machine Learning In Python", scikitlearn - https://scikit-learn.org/stable/

[10] Python Flask Server, Quickstart - https://flask.palletsprojects.com/en/1.1.x/quickstart/

[11] GridSearchCV, scikit learn - https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

[12] Kaggle Dataset - https://www.kaggle.com/datasets

[13] Xian Guang LI, Qi Ming LI, "The Application Of Data Mining Technology In Real Estate Market Prediction", irbnet.de, 2006

[14] Aswin Sivam Ravikumar, " Real Estate price Prediction Using Machine Learning", 2018

[15] Nihar Bhagat, Ankit Mohokar, Shreyash Mane, " House Price Forecasting Using Data Mining", 2016

[16] Wei Luo, "Guidelines for developing and reporting machine learning predictive models in biomedical research", December 2016

[17] Djamael Djenouri, "Machine Learning For Smart Building Applications: Review And Taxonomy", 2018

[18] lr-clr, linear classication, Medium - https://medium.com/@jouneidraza522/comparison-of-classification-algorithms- lr-dt-rf-svm-knn-6631493e300f

[19] K Fold Cross Validation, scikit learn - https://scikitlearn.org/stable/modules/generated/sklearn.model_selection.KFold.html

ISSN: 2233-7857IJFGCN

ISSN: 2233-7857IJFGCN