American Journal of Electronics & Communication

Madhulekha Hazra et. al., American Journal of Electronics & Communication, Vol. II (1), 13-15

ANALYSIS ON THE STATUS OF RENT OF HOUSING INDUSTRY

Madhulekha Hazra[1], Bikram Bhattacharya[2], Suryasish Sengupta[3], Rajesh Mandal[4], Poojarini Mitra[5], Kaustuv Bhattacharjee[6], Anirban Das[7]

Department of Computer Applications
University of Engineering & Management, Kolkata
Kolkata, India

[1]madhulekhahazra29@gmail.com, [2]bhattacharyabikram192@gmail.com ,[3]s.sengupta77777@gmail.com, [4]rajeshonline38@gmail.com, [5]poojarini.mitra@uem.edu.in, [6]kaustuv.bhattacharjee@uem.edu.in, [7]anirban-das@live.com

***Abstract* - Machine learning has been playing an active role in the past few years in several applications like image detection, spam reorganization, recommending products and in medical fields. Present machine learning algorithm helps us in enhancing security alerts, ensuring public safety and improve medical enhancements. Determining the sale price of the house is very important nowadays as the price of the land and price of the house increases every year. So our future generation needs a simple technique to predict the house price in future. The price of house helps the buyer to know the cost price of the house and also the right time to buy it. The right price of the house helps the customer to elect the house and go for the bidding of that house. There are several factors that affect the price of the house such as the physical condition, location, landmark etc. Our result exhibit that our approach to the issue needs to be successful, and can process predictions that would be comparative with other house rent prediction models. This paper uses linear regression technique to predict the house price.**

Index Terms* -. *Linear Regression, prediction analysis, machine learning

I. INTRODUCTION

Real property is not only a man's basic need, but it also represents a person's wealth and prestige today. Because their property values do not decline rapidly, investment in real estate generally seems to be profitable. Changes in the price of real estate can affect various investors in households, bankers, policymakers.

Investment in the real estate sector appears to be an attractive investment choice. Predicting the value of the immovable property is therefore, an essential economic index. This paper brings the latest research on regression techniques that can be used for house prediction, such as linear regression. As the initial house price prediction was challenging and requires some best method to get an accurate prediction. Data quality is a crucial factor to predict house prices and missing features are a challenging aspect to handle in machine learning models, let alone the house prediction model. Therefore, feature engineering becomes an essential method for creating models which will give better accuracy. In general, the value of the property increases over time, and its value must be calculated. During the sale of property or while applying for the loan and the property's marketability, this valued value is required. The professional evaluators determine these valued values. However, the disadvantage of this practice is that these evaluators could be biased because buyers, sellers, or mortgages have bestowed interest. We, therefore, need an automated model of prediction that can help to predict property values without bias. This automated model can help first-time buyers and less experienced customers to see if property rates are overrated or underrated.

Kaggle organizes a dataset "housing" of USA. The significant factors are Average Area Income, Average Area House Age, Average Area Number of Rooms, Average Area Number of Bedrooms, Area Population, Price, Address. The dataset contains various missing critical features which can degrade the model performance to predict house prices.

II. DESIGN & ANALYSIS

The dataset contains Numeric Variables:
X is a dependent variable, and y is an independent variable.
X variable contains average age income, Average area house age, Average area number of rooms, Average area number of bedrooms, Area population.
Y variable contains Price.

A. Linear regression

The Simple linear regression statistical method allows us to summarize and study the relationship between two continuous quantitative variables.

One variable, denoted x, is regarded as the predictor, explanatory, or independent variable.

The other variable, denoted y, is regarded as the response, outcome, or dependent variable.

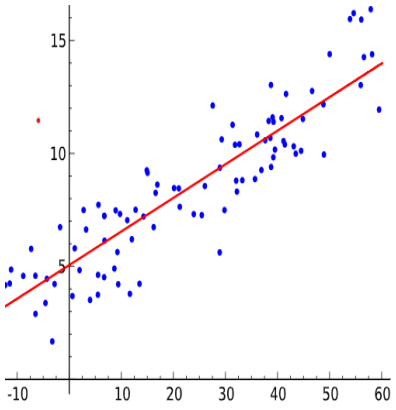

Fig. 1 Linear Regression

B. Multiple Linear Regression

Multiple linear regression (MLR), which is also known as multiple regression. It is used to predict outcomes on variables using statistical approaches. The purpose of MLR is to predict and display the outcome in a model based on the relation between dependant and independent variables.

In essence, multiple regression is the extension of ordinary least-squares (OLS) regression that involves more than one explanatory variable.

Multiple linear regression is the expansion of ordinary least-squares (OLS) regression. Unlike OLS it uses more than one independent variable.

$yi=\beta\ 0+\beta\ 1xi1+\beta\ 2xi2+...+\beta\ pxip+\epsilon$

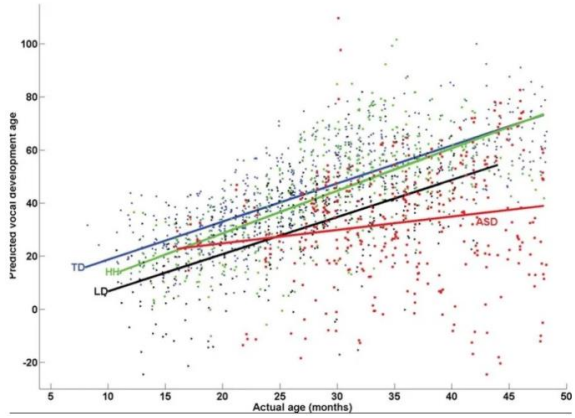

Fig. 2 Multiple Linear Regression

III. METHODOLOGY

Linear Regression is a supervised machine learning algorithm where the predicted output is continuous and has a constant slope. Instead of classifying the values into categories it predicts them within a continual range.

Simple regression

y= MX +by= MX+b

Multivariable regression

f(x,y,z)=w1x+w2y+w3z

The dataset, which is a CSV file, contains 5000 rows and seven columns.

The columns are-

1 Avg. Area Income
2 Avg. Area House Age
3 Avg. Area Number of Rooms
4 Avg. Area Number of Bedrooms
5 Area Population
6 Price
7 Address

The following libraries are used as shown in Fig.3.

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

%matplotlib inline
```

Fig. 3 Python libraries used,

A glimpse of a dataset is shown in Fig. 4

```
df=pd.read_csv("/content/drive/MyDrive/USA_Housing.csv")
df.head()
```

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Address
0	79545.458574	5.682861	7.009188	4.09	23086.800503	1.059034e+06	208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
1	79248.642455	6.002900	6.730821	3.09	40173.072174	1.505891e+06	188 Johnson Views Suite 079\nLake Kathleen, CA...
2	61287.067179	5.865890	8.512727	5.13	36882.159400	1.058988e+06	9127 Elizabeth Stravenue\nDanielltown, WI 06482...
3	63345.240046	7.188236	5.586729	3.26	34310.242831	1.260617e+06	USS Barnett\nFPO AP 44820
4	59982.197226	5.040555	7.839388	4.23	26354.109472	6.309435e+05	USNS Raymond\nFPO AE 09386

Fig. 4 Glimpses of the dataset

The describe() method is used for calculating and displaying statistical data like count, min, max, std etc. of a DataFrame or series of numeric values. It analysis the DataFrame series of mixed data types and also individual numeric and object series.

```
df.describe()
```

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5.000000e+03
mean	68583.108984	5.977222	6.987792	3.981330	36163.516039	1.232073e+06
std	10657.991214	0.991456	1.005833	1.234137	9925.650114	3.531176e+05
min	17796.631190	2.644304	3.236194	2.000000	172.610686	1.593866e+04
25%	61480.562388	5.322283	6.299250	3.140000	29403.928702	9.975771e+05
50%	68804.286404	5.970429	7.002902	4.050000	36199.406689	1.232669e+06
75%	75783.338666	6.650808	7.665871	4.490000	42861.290769	1.471210e+06
max	107701.748378	9.519088	10.759588	6.500000	69621.713378	2.469066e+06

Fig. 5 describe the method to show data.

X is then taken as a dependent variable and y as an independent variable.

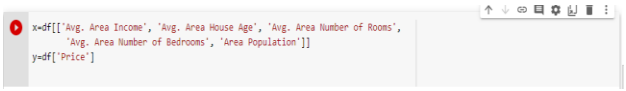

Fig. 6 Dependent and Independent variables assigned.

Data are split and trained to implement Linear Regression. Using the in-built function sklearn train_test_split() the data set is split into two parts. One part is called the training set, which is used to fit and train the model. Another part is called the testing set which is used to check the evaluation on the final model.

We then import LinearRegression from sklearn.linear_model and create an object of it and then run the LinearRegression() function. LR.fit is then performed.

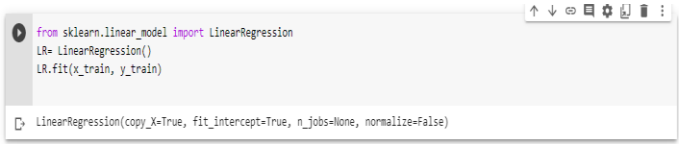

Fig. 7 Linear Regression method implementation

IV. RESULTS AND DISCUSSION

The model is evaluated ,and coefficients are checked.

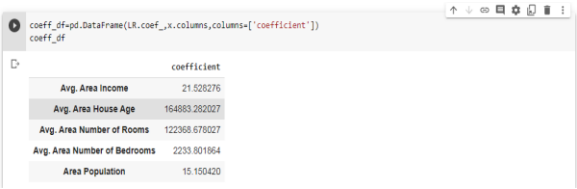

Fig. 8 Model evaluation and coefficient checking

The coefficient of data canvases:

1.) With other fields constant, 1 unit raise in Avg. Area Income has reflected a raise of $21.52.

2.) With fields constant, 1 unit raise in Avg. Area House Age has reflected a raise of $164883.28.

3.) With fields constant, 1 unit raise in Avg. Area Number of Rooms has reflected a raise of $122368.67.

4.) With fields constant, 1 unit raise in Avg. Area Number of Bedrooms has reflected a raise of $2233.80.

5.)With fields constant, 1 unit raise in Area Population has reflected a raise of $15.15.

The data plotting reflects the prediction in Fig.9

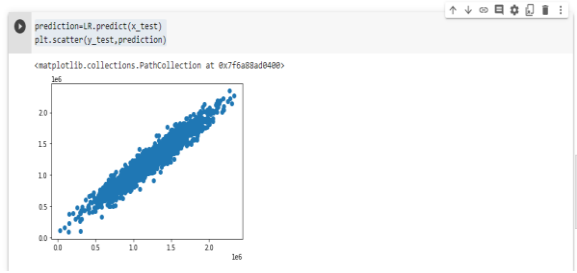

Fig. 9 Data plotting and prediction

The accuracy of the model means the dependent variable Y has a linear relationship to the independent variable X. To check this, we have to guarantee it that the resulted scatterplots using the X and Y coordinates are linear and that the remaining plot shows some incidental pattern.

IV. CONCLUSION

Most value is removed by roofing a home with clay tiles. Interestingly, being close to a park or other outdoor feature also lowers the home's value. Alternatively, the value is increased by a few neighborhoods. Generally, property values rise over time, and its appraised value needs to be calculated. We have mentioned the step-by-step procedure to analyze the dataset and find the correlation between the parameters. Thus we can select the parameters which are not correlated to each other and are independent. These feature sets were then given as an input to four algorithms, and the (.csv) file was generated consisting of predicted house prices. This research work can be useful for the development of applications for various respective cities.

To get the result, the input of four algorithms and a .csv file is generated representing predicted house prices. This research work is important as one can use it for the development of applications and accessing them from different cities.

Herby concluding that this research works on spatial statistics and the application can work with a large number of data.

REFERENCES

[1] Jakob A. Damon, Fabio Sigrist, Reinhard Furrer, Maximum likelihood estimation of spatially varying coefficient models for large data with an application to real estate price prediction, Spatial Statistics, Volume 41,2021

[2] B. Prashanth, Mruthyunjaya Mendu, Ravikumar Thallapalli,Cloud based Machine learning with advanced predictive Analytics using Google Colaboratory, Materials Today: Proceedings, 2021,ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.01.800.

[3] Quang Truong, Minh Nguyen, Hy Dang, Bo Mei,Housing Price Prediction via Improved Machine Learning Techniques, Procedia Computer Science, Volume174,2020,Pages433-442,1877-0509,https://doi.org/10.1016/j.procs.2020.06.111.

[4] Lily Shen, Stephen Ross, Information value of property description: A Machine learning approach, Journal of Urban Economics, Volume 121,2021,103299,ISSN00941190,https://doi.org/10.1016/j.jue.2020.103299.

[5] Khalid K. Al-jabery, Tayo Obafemi-Ajayi, Gayla R. Olbricht, Donald C. Wunsch II, 9 - Data analysis and machine learning tools in MATLAB and Python, Editor(s): Khalid K. Al-jabery, Tayo Obafemi-Ajayi, Gayla R. Olbricht, Donald C. Wunsch II, Computational Learning Approaches to Data Analytics in Biomedical Applications, Academic Press, 2020, Pages 231-290, ISBN 9780128144824,https://doi.org/10.1016/B978-0-12-814482-4.00009-7.

[6] Shui-xia Chen, Xiao-kang Wang, Hong-yu Zhang, Jian-qiang Wang, Customer purchase prediction from the perspective of imbalanced data: A machine learning framework based on factorization machine,Expert Systems with Applications,Volume 173, 2021, 114756, ISSN0957-4174, https://doi.org/10.1016/j.eswa.2021.114756

[7] Yuhao Kang, Fan Zhang, Wenzhe Peng, Song Gao, Jinmeng Rao, Fabio Duarte, Carlo Ratti, Understanding house price appreciation using multi-source big geo-data and machine learning, Land Use Policy, 2020, 104919,ISSN0264-8377,https://doi.org/10.1016/j.landusepol.2020.104919.

[8] https://scikit-learn.org/stable/modules/linear_model.html

[9] http://www.stat.yale.edu/Courses/1997-98/101/linreg.html

[10] http://home.iitk.ac.in/~shalab/regression/Chapter2-Regression SimpleLinearRegressionAnalysis.pdf.