Utilization Of Machine Learning Models In Real Estate House Price Prediction

Anurag Sinha
Department of computer science, Student, Amity University Jharkhand
Ranchi, Jharkhand(India), 834001
anuragsinha257@gmail.com

***Abstract*- Machine learning participate a significant role in every single area of technology as per the today's scenario. Even I can Say every phase of our lives is surrounded by the implementation of new era technologies such as Hospitality management, Railway, Transportation, Health care, Industry**
And so on. Machine learning has been employed for many sectors since past decades like image processing, pattern recognition, medical diagnosis, and predictive analysis, product recommendation. House prices changes every year, so it is mandatory for a structure to foresee house prices in the future. House price prediction can help in fixing and thereby predicting house prices and customer can evaluate it. Our intension is to predict house prices using several machine learning techniques. House price of particular location does depends on various factors like lotsize, bedrooms, bathrooms, location, drawing room, material used in house , interiors, parking area and mainly on square feet per area.
Our intension behind proposing this paper is to employ different machine learning techniques for predicting the price based on these metrics. The algorithm used in this analysis is Data refining, OLS regression, Classification, Clustering, correlation matrix.

Keywords: Real estate, machine learning, classification, Clustering, OLS regression, prediction model

I. INTRODUCTION

Real Estate Property can't be taken only as important need of society but today it also represents the status and reputation of an individual. Venture in real estate usually seems to be gainful because their possessions values do not turn down rapidly. Transformation in the real estate price can control a mixture of conjugal sponsors, bankers, policy makers and so on. Venture in real estate region seems to be an eye-catching alternative for the investments. Investment is commerce that most people are interested in this globalization age. There are numerous often used for investment, for example, gold, stocks and property.

Machine learning aims at developing self-learning algorithms using datasets, such that machine can be enabled to project future activity based on the past data. It helps organization in easily identifying and predicting the trends and patterns. It also facilitates the managers, executive and
Analysts make use of its models in making decision in efficient manner. It helps the organization to adopt environmental without human intervention

Machine learning is used as an algorithm for building model and thereby using that model for predicting new a data set. The prominent difference of using conventional algorithm that result is oriented with the input data rather than focusing on a chain of different instructions set. Supervised learning is based on building a model based on labeled data set whereas Unsupervised learning is totally oriented with unlabelled data set. There are several machine learning algorithm are regression, classification, clustering, SVM, neural network, deep learning and so on. It is crucial to predict the definite outcome out of model which is based on feature extraction. So far the matter of house prediction is concern it is determined by following metrics:-

- location,
- dimension,
- house type,
- city,
- country,
- tax rules,
- economic cycle,
- population movement,
- Interest rate, and a lot of other aspects which could control demand and supply.

II. LITERATURE REVIEW

In our literature survey we have investigated various researches on this particular domain some of them are as follows:-

Author Vargas and Silva showed a comparative study about lodging cost are considered as a main pillar in the grit of the business sector. When the bangs and all other attributes when increase in lodging division such as development and work then these must have a balance with accumulation method, and thus resulting the cost and value of house increasing upwards. Consequently, all

these are collaborated and result in withdrawal phase, drop in the private money cause the interest rate of house increase drastically. By the perceived the house cost based on these model influence the cost of house in real life and householders are pressurized to diminish their actual hose cost. The total agreement will only be possible to achieve when the declination in rate will compliment the bargain value of the house. Moreover, throughout withdrawal and subsidence factual house costs fall quickly Likewise general inflationary patterns reduces the accurate house costs much with closed perceived costs.

The land costs are predicted with a new set of factors with a different method. Also we predicted the reimbursement for the agreement of the property. Mathematical relationships help us to understand many portions of everyday life. When such relationships are communicated with exact numbers, we gain additional clearness Regression is concerned with specifying the relationship between a solitary numeric dependent variable and one or more numeric independent variables. House prices have tendency to increase very year
So there must be a concrete way of predicting the price effectively. House price prediction helps the developer forecast the house price on a genuine cost range and clients too to manage to buy house when needed.

III. PROBLEM IDENTIFICATION AND DATA COLLECTION

For our research we have collected data for real estate from
Ranchi capital of Jharkhand .we have employed data and
These metrics of real estate of Ranchi zone, various factor
Like lotsize, bedrooms, bathrooms, location, drawing room,
Material used in house, interiors, parking area and mainly On square feet per area. Further use these data for building Data model for predicting the house price after data Analysis. We gathered the real estate data from every zone and sub divisional area of Ranchi, Jharkhand. These are the labeled and structured data set presentation.

IV. MACHINE LEARNING TECHNIQUE FOR BUILDING MODEL

TABLE 1- PRICE FACTORS

Avg. Price / Sqft	area	lotsize	bedrooms	bathroom
₹2,371	rnc	850	3	1
₹3,761	rnc	4000	4	2
₹3,724	rnc	3060	3	1
₹3,724	rnc	6650	3	1
₹3,097	rnc	6360	3	2
₹4,093	rnc	7383	6	3
₹3,396	rnc	6734	5	2
₹3,396	rnc	9866	4	1
₹3,396	rnc	7888	4	1

DATA REFINING

Data refinery brings wide-ranging multifaceted events processing functionality that make sure that incident data is quickly and efficiently donate to use, to advance business routine and customer experience. Data refinery participate an imperative role in supporting the digital customer journey. It provides a cover of intellectual fast data to facilitate real-time decisions and actions. It gathers, unifies and refines data from numerous sources into valuable inputs from multiple sources.

OLS REGRESSION

This process minimizes the sum of squared perpendicular distances between the observed responses in the dataset and the responses predicted by the linear rough calculation The resulting estimator can be expressed by a straightforward formula, especially in the case of a single repressor on the right-hand side.

Linear regression is often used to foresee outputs clue for new samples. This enables the organizations to distinguish the value of the model and for the prediction before they go further on and use it for predictive use.

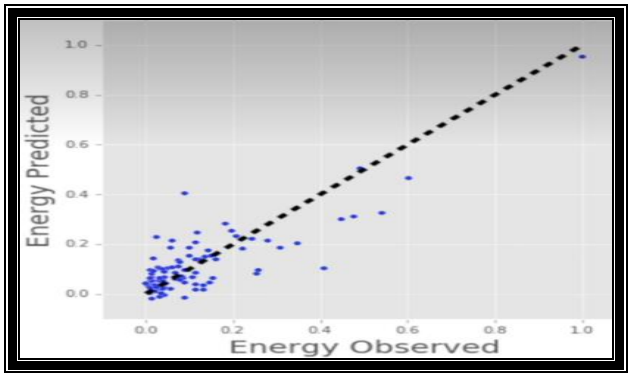


Figure 1- OLS regression

TABLE 2- COST DRIVERS

no of floor	driveway	fullbaseme	garage	balcony
2	yes	yes	yes	no
2	yes	no	yes	no
1	no	yes	yes	yes
2	no	no	yes	yes
3	no	yes	yes	no
2	yes	no	no	yes
1	yes	yes	no	no
3	yes	no	no	yes
2	ys	yes	yes	no

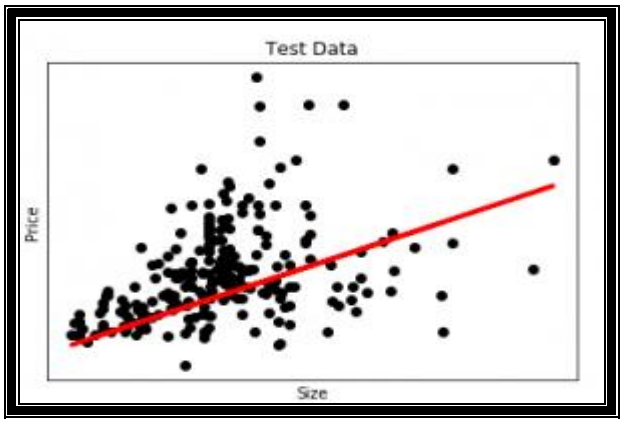


Figure 2- test data set of real estate

CLASSIFICATION

Classification is a type of supervised learning. It specifies the group to which data elements fit in to and is best used when the output has predetermined and separate values. It calculates a division for an input variable as well. Classification is a progression of categorizing a specified set of data into classes; it can be performed on both structured and unstructured data. The procedure starts with predicting the class of given data points. The classes are often referred to as objective, label or categories. The classification predictive modeling is the job of approximating the mapping purpose from input variables to separate output variables. The major purpose is to recognize which class/category the new data will fall into.

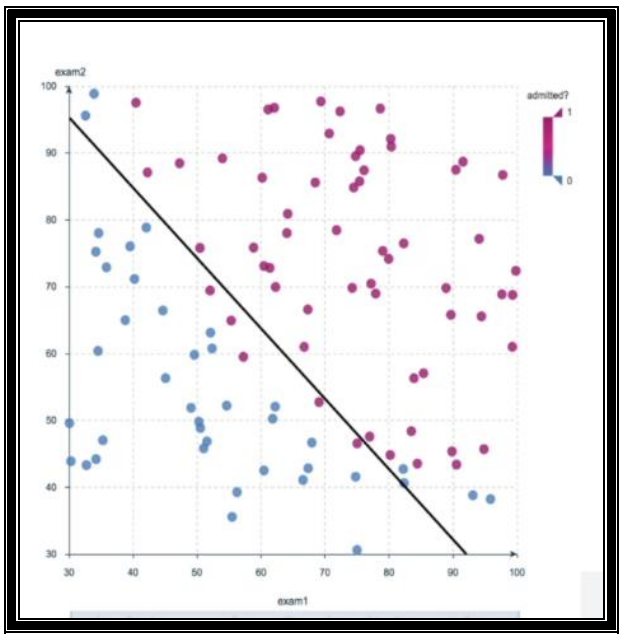
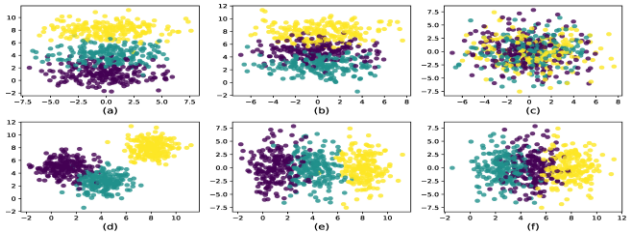


Figure 3- classification

CLUSTERING

Clustering can be defined as it is having task of separating the occupiers or data points into group of numbers such that data points in the corresponding groups are more as



good as to other data points in the identical group than those in other groups. In simple words, the intension is to separate out the groups with related traits and allocate them into clusters.

Table 3: A comparison of machine learning house price predictions (2016-2018)

Paper	Linear Regression	Random Forest	Decision Tree	Ridge	Lasso
[20]		✓	✓		
[24]	✓	✓	✓		✓
[25]		✓			
[26]	✓	✓			
[27]		✓		✓	✓
[28]		✓	✓		
[29]		✓			
[30]				✓	✓
[31]	✓	✓			
[32]	✓	✓			

V. METHODOLOGY & RESULTS

Dataset that which is employed in this technique is taken from the list of real estate house price list based on the several locations of Ranchi zone. As the scope of this paper to predict the house cost thus for the sake of algorithm several metrics are used as for feature extraction and these variables are called feature data set. Table 4 shows the combination of different set of features which are used to build prediction model. This proposed approach comprises
19 different attribute or feature set as autonomous variables for predicting house prices.

Features selection is an imperative footstep of machine learning prediction. In this paper, features selection is divided into four groups. First collection used all the self-determining parameters in the training dataset. It is a blend of variables with very weak, weak and strong relationships on the related variable sale price. In this paper, the point of association is defined as Strong if the coefficient correlation value is between some range and astute if the value is between (0.3 to 0.5). Otherwise, weak level is between (0.2 to 0.29) and very weak level is between 0.1 to 1.

Features	Description
Selling price	Dispose price/sqf (RM)
Buying price	Transaction price/sqf (RM)
Floor	Floor
GC	Green certificate
MFA	Main floor area
Bed	Number of bedrooms
Distance	Distance to CBD
BC	Building category
Ownership	Own
CA	Category area
AC	Area classification
Floor	Floor
BC	Building category
CLASS	Building classification
Bed	Number of bedroom
Age	Age of the building
Buy	Buyers
Sell	Seller

Figure and table 4- features

Table 5-OLS regression result

OLS Regression Results

Dep. Variable:	price	R-squared (uncentered):	0.956
Model:	OLS	Adj. R-squared (uncentered):	0.956
Method:	Least Squares	F-statistic:	1067.
Date:	Mon, 15 Jul 2019	Prob (F-statistic):	0.00
Time:	06:03:17	Log-Likelihood:	-6034.8
No. Observations:	546	AIC:	1.209e+04
Df Residuals:	535	BIC:	1.214e+04
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Lotsize	3 .4431	0.339	10.144	0.000	2.776	4.110
bedrooms	1095.9263	842.938	1.300	0.194	-559.947	2751.800
bathrms	1.402e+04	1466.301	9.561	0.000	1.11e+04	1.69e+04
stories	6526.5732	925.283	7.054	0.000	4708.940	8344.206
driveway	5665.6447	1854.971	3.054	0.002	2021.724	9309.565
recroom	4659.4642	1896.548	2.457	0.014	933.870	8385.059
fullbase	5306.1054	1583.810	3.350	0.001	2194.856	8417.355
gashw	1.285e+04	3218.757	3.993	0.000	6529.985	1.92e+04
airco	1.28e+04	1549.330	8.260	0.000	9754.655	1.58e+04
garagepl	4379.7318	833.106	5.257	0.000	2743.173	6016.291
prefarea	9561.2358	1661.849	5.753	0.000	6296.687	1.28e+04

Omnibus:	101.942	Durbin-Watson:	1.576
Prob(Omnibus):	0.000	Jarque-Bera (JB):	279.382
Skew:	0.915	Prob(JB):	2.15e-61
Kurtosis:	5.988	Cond. No.	2.74e+04

Regression uses random smallest amount squares technique to the modified data for precise prediction of house prices based on the refined dataset variables.

The regression coefficient represented by coef in the fig shows the change in the dependent variable “house price” resulting from one unit change in the particular variable, all other variables being held constant

As a result the location of the house, along with the amenities are highly influenced

Subsequently, the variables and their correlation matrix are described below:-

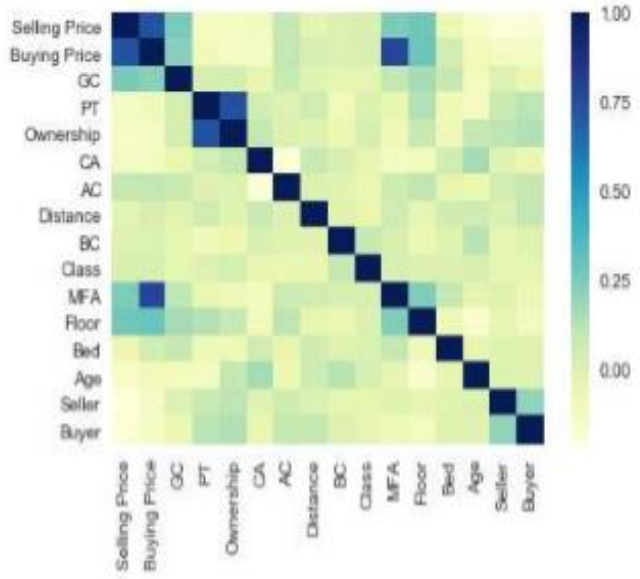


Figure 5- Correlation level of all features

Figure 6 - Correlation level of features with values

Figure 7 - Correlation level of features with values between {0.1 to 0.19}

Only the purchasing cost variable is found to have a strapping connection with the selling prices with coefficient value 0.73, as presented in the following

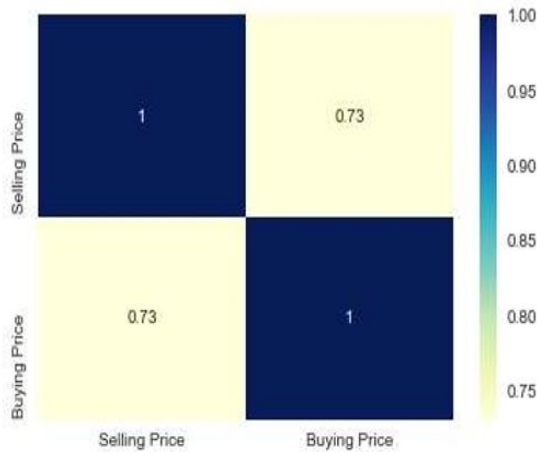


Figure 8 -The correlation level between selling price and buying price

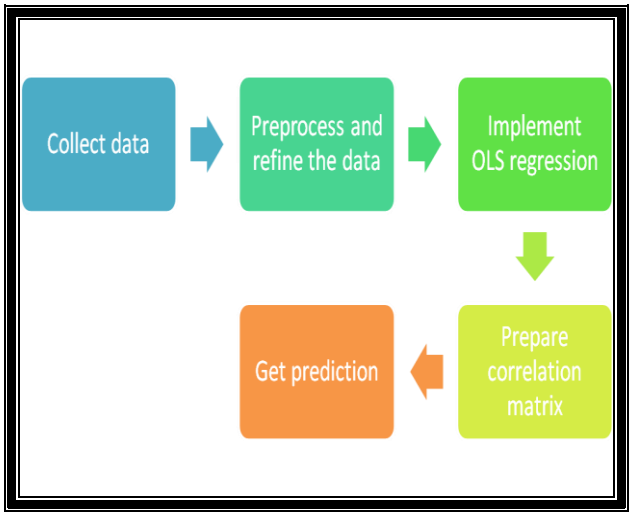


Figure 9- System Work Flow

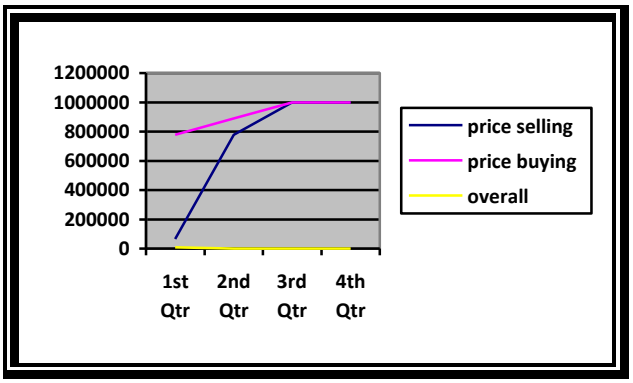


Figure 10 – prediction in increasing slope of price

VI. CONCLUSION

Taking the data set for houses and considering its various attributes, the prices for houses have been predicted by employing machine learning methods of regression for predicting the price of ESTATE USING RIOR DATA, AND CLUSTERING FOR INSPECRING THE QUALITY OF THE SOLUTION OR OUTPUT.

Ml driven predictions are easily comprehensible and significant from a data analysis of point. When correctly implemented a high rate of accuracy can be achieved, and thus ML techniques find applications across a wide range of fields.

VII. REFERENCES

[1] G. Naga Satish, Ch. V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu"House Price Prediction Using Machine Learning" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-9, July 2019

[2] Bruno Klaus de Aquino Afonso1 , Luckeciano Carvalho Melo2 , Willian Dihanster Gomes de Oliveira1 , Samuel Bruno da Silva Sousa1 , Lilian Berton1 "Housing Prices Prediction with a Deep Learning and Random Forest Ensemble"1https://www.kaggle.com/c/data-science-challenge-at-eef-2019/

[3] Ayush Varma ; Abhijit Sarma ; Sagar Doshi ; Rohini Nair" House Price Prediction Using Machine Learning and Neural Networks" 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)

[4] Jingyi Mu,1 Fang Wu,2 and Aihua Zhang "Housing Value Forecasting Based on Machine Learning Methods" Hindawi Publishing Corporation Abstract and Applied Analysis Volume 2014, Article ID 648047, 7 pages http://dx.doi.org/10.1155/2014/648047.

[5] NEELAM SHINDE, 2KIRAN GAWANDE "VALUATION OF HOUSE PRICES USING PREDICTIVE TECHNIQUE" International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835 Volume-5, Issue-6, Jun.-2018

[6] Prof. Pradnya Patil "House Price Prediction Using Machine Learning and RPA" International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

[7] Hromada, Eduard "mapping of real estateprices using data mining techniques" procedia engineering123(2015) 233-240.

[8] S. Sirmans, D. Macpherson, and E. Zietz, "The composition of Hedonic pricing models," J. Real Estate Lit., 13(1), 2005, pp. 1–44.

[9] M. Sasaki and K. Yamamoto, "Hedonic price function for residential district that specialize in the explanations for residential preferences in Japanese metropolitan areas," J. Risk Financ. Manag., 11(3), 2018, pp. 2–18.

[10] C. K. Wing, S. K. Wong, and L. W. C. Lai, "a Hedonic price modelling of environmental attributes: A review of the literature and a Hong Kong case study," Underst. Implement. Sustain. Dev., 2002, pp. 87–110.