iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 11 **Issue:** VII **Month of publication:** July 2023

**DOI:** https://doi.org/10.22214/ijraset.2023.54739

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 11 Issue VII Jul 2023- Available at www.ijraset.com*

# House Price Prediction Using Machine Learning

Dipanshu[1], Tarun[2], Harsh, Keshav[3]
*Branch of Computing Science and Engineering, Chandigarh University, Mohali, India*

***Abstract:*** ***The actual estate marketplace is a hero among those most worried about valuation, and it is still evolving. It's still one of the best places to put AI's thoughts to work, as it's the most effective way to improve and predict costs with high precision. The purpose of this study is to present a market value assumption for land property. This method detects a beginning cost for a property based on topographical factors. Future expenses will be norm aliased by isolating prior market models and worth ranges, as well as upcoming forms of advancement. With a Decision tree regressor, this evaluation can estimate house costs in Mumbai. It will help clients with putting properties into a domain without touching towards a seller. The delayed consequence of this assessment showed that the Decision tree regressor gives an accuracy of 89%.***
***Keywords: Multiple linear regressions, House Price Prediction, Machine Learning.***

## I. INTRODUCTION

Every single relationship in the current land business is working helpfully to achieve an upper hand over elective players. It is necessary to improve collaboration for a typical individual while achieving the greatest results. This research offers a system that uses backslide AI to forecast housing expenses. If you will sell a house, you want to see what retail cost to put on it. Moreover, a PC assessment can give you a precise measure. This backslide model is collected not only for expecting the expense of the house which is arranged accessible to be bought at this point also for houses that are a work in progress. The real estate market is a hero among the most preoccupied with valuing and continues to evolve.

Inversion is an AI device that urges you to do what is generally anticipated of you by taking - from the current quantifiable data - the association between the control boundary and the number of various boundaries. As indicated by this definition, the expense of a house relies upon the boundaries, for instance, the number of rooms, convenience, convenience, etc. In the impossible occasion that we utilize extortion that observes these cutoff points, we can compute lodging costs in a specific region of the world.

The objective part in The cost of the land property is included in this suggested model, as are the independent components: no. of rooms, no. of washrooms, cover locale, created area, the floor, age of the property, postal region, degree, and longitude of the property. Other than those of the referred to features, which stay generally expected at expecting the house costs, We've incorporated two different components: air quality and noise pollution. These features give a significant responsibility towards predicting property costs since the higher potential gains of these components will incite a reduction in house costs.

The whole execution is done using the python programming language. For the advancement of the farsighted model, The "Scikit-learn" AI library's decision tree regressor is employed. Structure Search CV supports with finding the best max-significance and impetus for fostering the result tree. After the pre-planned model is complete, the UI is created using Flask (a Python framework).

## II. LITERATURE SURVEY

### *A. Real Estate Price Prediction with Regression and Classification*

This paper, house costs are anticipated utilizing logical factors that cover numerous parts of private houses. House costs are anticipated with different relapse procedures Lasso, Ridge, SVM relapse, and Random Forest are just a few examples. According to this article, the best-performing model for a relapse issue is SVR with Gaussian bit, which has an RMSE of 0.5271. in any case, representation for SVR was troublesome because of its high dimensionality. As per its examination, residing region rectangular feet, the material of the rooftop, and locality have the best measurable importance in foreseeing a family's deal cost.

### *B. An SVR Based Forecasting Approach for Real Estate Price Prediction*

The help vector machine (SVM) has been applied successfully to order, bunch, and gauge. The support vector relapse (SVR) method is proposed in this paper to estimate land costs in China. The point of this newspaper was to analyze the attainability of SVR in land cost expectation. The test results were determined because of the mean outright mistake (MAE), the mean outright rate mistake (MAPE) and the root mean squared error (RMSE), as well as the SVR-based method was an effective device to gauge land costs.

©IJRASET: All Rights are Reserved 

*C. House Price Prediction Using Machine Learning Algorithms*

The demand for renting and owning homes has increased as a result of increased urbanisation. As a result, figuring out a better technique to compute property prices that truly represent market prices has become a popular issue. The research focuses on applying machine learning methods such as simple linear regression (SLR), multiple linear regression (MLR), and neural networks to properly determine the house price (NN). The algorithm with the lowest Mean Square Error (MSE) is picked as the best for estimating the price of a property. This will assist both sellers and buyers in determining the optimal price for a home.

*D. House Price Prediction Using Machine Learning and Neural Networks*

In this paper means to make assessments in light of each essential boundary that is considered while deciding the cost. This model involved different relapse strategies in its is pathway, and the outcomes are still up in the air through single strategy instead, it's the weighted average of different procedures to give the most dependable outcomes. The outcomes demonstrated that this method yields the least mistake and most extreme precision than separate calculations useful.

*E. Vision-based real Estate Price Estimation*

The subject of automated calculation of market prices for properties has gotten a lot of attention with the development of online real estate database businesses like Zillow, Trulia, and Redfin. Several real estate websites use a proprietary methodology to offer such estimations. Although these predictions are frequently close to the actual sale prices, they can be quite incorrect in other circumstances. The inside and external look of a home are important aspects that are not taken into account when creating computerised value estimations. The influence of a house's aesthetic qualities on its market value is examined in this research. We build a method for predicting the luxury level of real estate photographs by using deep convolutional neural networks on a huge dataset of photos of property interiors and exteriors. We also create a new framework for automatic value evaluation that incorporates the above photographs as well as property parameters such as size, proposed price, and number of bedrooms. Finally, we demonstrate that our proposed technique for price estimation surpasses Zillow's estimations by applying it to a fresh dataset of real estate photographs and information.

## III. RELATED WORK

The worth of a specific property relies on the design facilities encompassing the property. Recently, a few scholars' increases for seeing the best properties for the clients showed with different advances. Raghunandan [1] alluded to the major information mining considerations of how it limits and supports calculations with a definitive goal of suspicion. The essential part is which AI assessment is the most fitting for expecting the house cost. As frequently as conceivable the locale's normal circumstances wrap up what sort of huge worth we can expect for various kinds of houses, Manjula [2] presents different basic elements to use while expecting property costs with extraordinary precision utilizing a loss the faith model. A. Varma [3] organized a framework that utilized consistent neighborhood information to get exact genuine valuations utilizing Google maps. Experts furthermore showed that there exist associations between physical appearance and non-physical appearance qualities like bad behavior bits of knowledge, dwelling costs, people thickness, etc of a city. For instance, Using Visual Basics to Expect Non-Visual City Attributes" [4], usages visual qualities to anticipate the arrangement cost of the stuff. used gathering and backslide estimations. As shown by examination, living district square feet, roof content, and neighborhood have the best quantifiable importance in evaluating the selling cost for a home. Besides assumption examination The PCA approach will help you enhance your results. focused on several estimations like brain association and Radiated reason pragmatic (RBF) brain associations. The RBF and BPN models are well-known with recognizing the differentiation between the house estimation document like Cathy and sinny cost record and jumbled relationship ability to distinguish the macroeconomic assessment. Nihar Bhagat, Ankit Mohokar, Shreyash Mane (2016) [7] read up direct backslide computations for the figure of the houses. The goal of the paper is to predict the useful expense of land for clients concerning their monetary plans and needs. Assessment of past market examples and worth arrives voluntarily expect future house assessing.

## IV. SYSTEM DESIGN AND ARCHITECTURE

*1) Stage 1: Collection of data*

Information handling strategies and cycles are various. We gathered the information for Mumbai's land properties from different land sites. The information would have traits, for example, Location, cover region, developed region, age of the property, postal district, and so forth We should gather the quantitative information which is organized and ordered. Information assortment is required before any sort of AI research is completed. Dataset legitimacy is an unquestionable requirement in any case it is a waste of time to break down the information.

*2) Stage 2: Information Preprocessing*

Data preprocessing is the most well-known approach to cleaning our instructive file. There might be missing characteristics or irregularities in the dataset. Data cleansing can help with these issues. Expecting a variable to have a large number of missing attributes we drop persons characteristics else substitute them with the typical worth.

*3) Stage 3: The model's education*

We should train the model first since the data is separated into two modules: a Training set and a Test set. The objective variable is joined by the readiness set. The layout of educational assortment is computed using a decision tree regressor. A backslide model is collected as a tree structure by the Decision tree.

*4) Stage 4: Testing and Integrating with UI*

The test dataset is fed into the pre-programmed model, and home expenses are predicted. The front end, which includes Flask in Python, is then designed using the pre-arranged model.

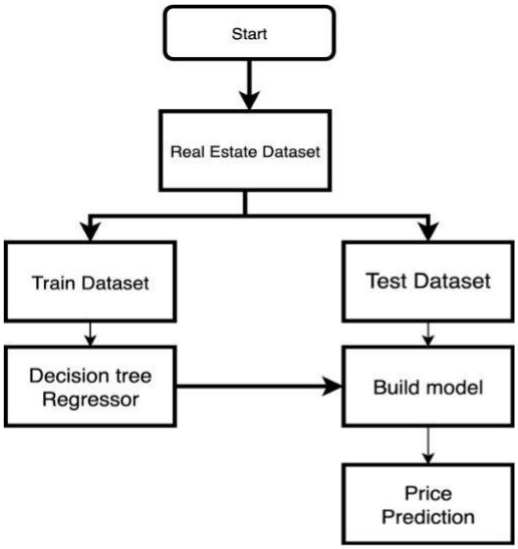



Figure 1 depicts the general development process.

## V. METHODOLOGY

*A. Studied Algorithms*

During the time spent encouraging this model, different backslide computations were thought about. Straight backslide, Multiple straight backslide, Decision Tree Regressor, and KNN are all examples of machine learning techniques. were attempted upon the planning dataset. In any case, the decision tree regressor gave the most raised accuracy to the extent that expecting the house costs. The decision to pick the computation particularly depends on the angles and the type of data in the data that was used For our dataset, the decision tree computation is the best option.

*B. Regressor For Decision Tree*

The decision tree regressor recognises quality components and trains a model like a tree to forecast data in the future to provide a massive result. The highest significance and minimum significance of a chart are gained by the decision tree regressor, which then separates the data as demonstrated by the system. Network Search CV is a strategy for overseeing limit tuning that will beneficially create and study a model for each mix of calculation limits exhibited in a cross-section. System Lookup In this calculation, CV is utilised to determine the optimal impetus for max-significance, which is then used to construct the decision tree.

*C. Flask Integration*

Right after building the model and giving the result, the accompanying stage is to do the consolidation with the UI, hence a cup is used. A carafe is a web structure. This implies that a carafe provides you with equipment, libraries, and headways that grant you to gather a web application. Flask is a framework for connecting Python models that is simple to learn.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

## VI. IMPLEMENTATION

### *A. Data Preprocessing*

For their missing traits, age and floor restrictions were addressed. In addition, the preparation dataset's goal attribute is removed. This is why the Pandas library is used. The objective characteristic's min, max, standard deviation, and mean were identified for the factual representation of the dataset. We divided the dataset into two parts: a preparation set (80%) and a test set (20%). (20 percent).

### *B. Max-profundity*

As previously said, system scan cv assists in determining the bush's maximum relevance. To visualise the various max-profundities and unpredictable execution, we utilised Matplotlib.

The visuals are as follows:

*1) Graph*

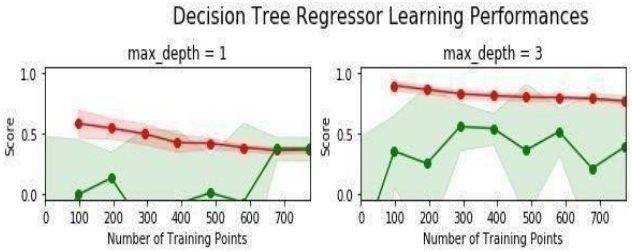



Fig. 2: Maximal Profundity Values are Tested (1)
(Score vs. Number of preparation points on pivot)

*2) Graph*

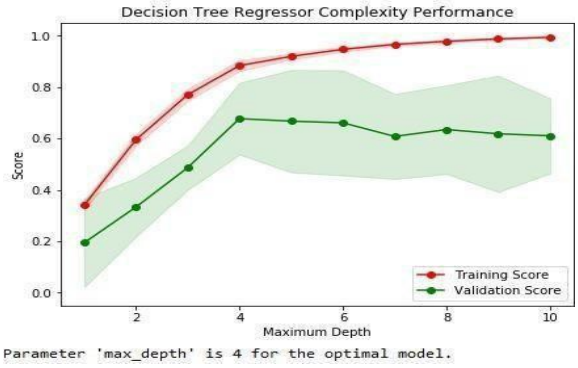



Fig. 3: Maximal Profundity Values are Tested (2)
(On pivot: number of points prepared vs. score)

*3) Graph*

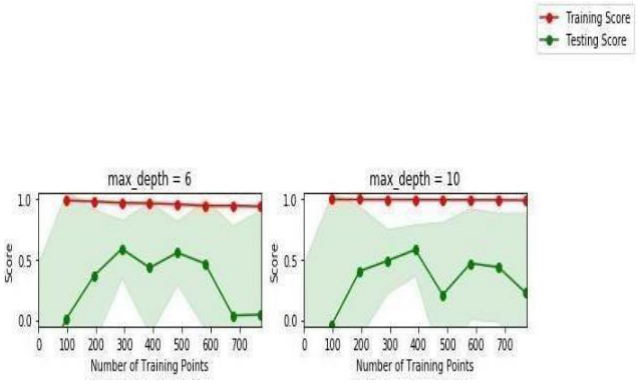



Fig 4. Max-profundity is an incentive for an ideal model

### *C. Fitting the Model*

The model is built using a Decision tree regressor from the Scikit-learn module. The test set results are predicted using the predict work.

## VII. OUTCOME

An accompanying displays the design of anticipated versus genuine costs with the precision of expectation:

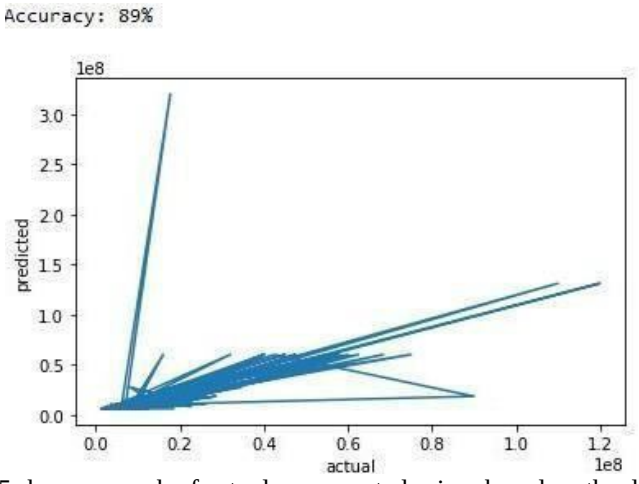



Figure 5 shows a graph of actual vs. expected prices based on the dataset.

The r2 score of the regression model is what determines accuracy.

## VIII. FUTURE SCOPE

Future work on this study could be separated into four fundamental regions to further develop the outcome even further. This should be possible by:-

*1)* The utilized pre-handling strategies truly do help in the forecast exactness. Nonetheless, exploring different avenues regarding various blends of pre-handling strategies to accomplish better expectation exactness.
*2)* Utilize the accessible elements and assuming they could be joined as binning highlights has shown that the information got moved along.
*3)* Preparing the datasets with various relapse strategies, for example, Elastic net relapse that consolidates both L1 and L2 standards. To grow the examination and check the execution.
*4)* The connection has shown the relationship in the neighborhood information. In this way, endeavoring to improve the neighborhood information is expected to cause rich with highlights that fluctuate and can give a solid connection relationship.

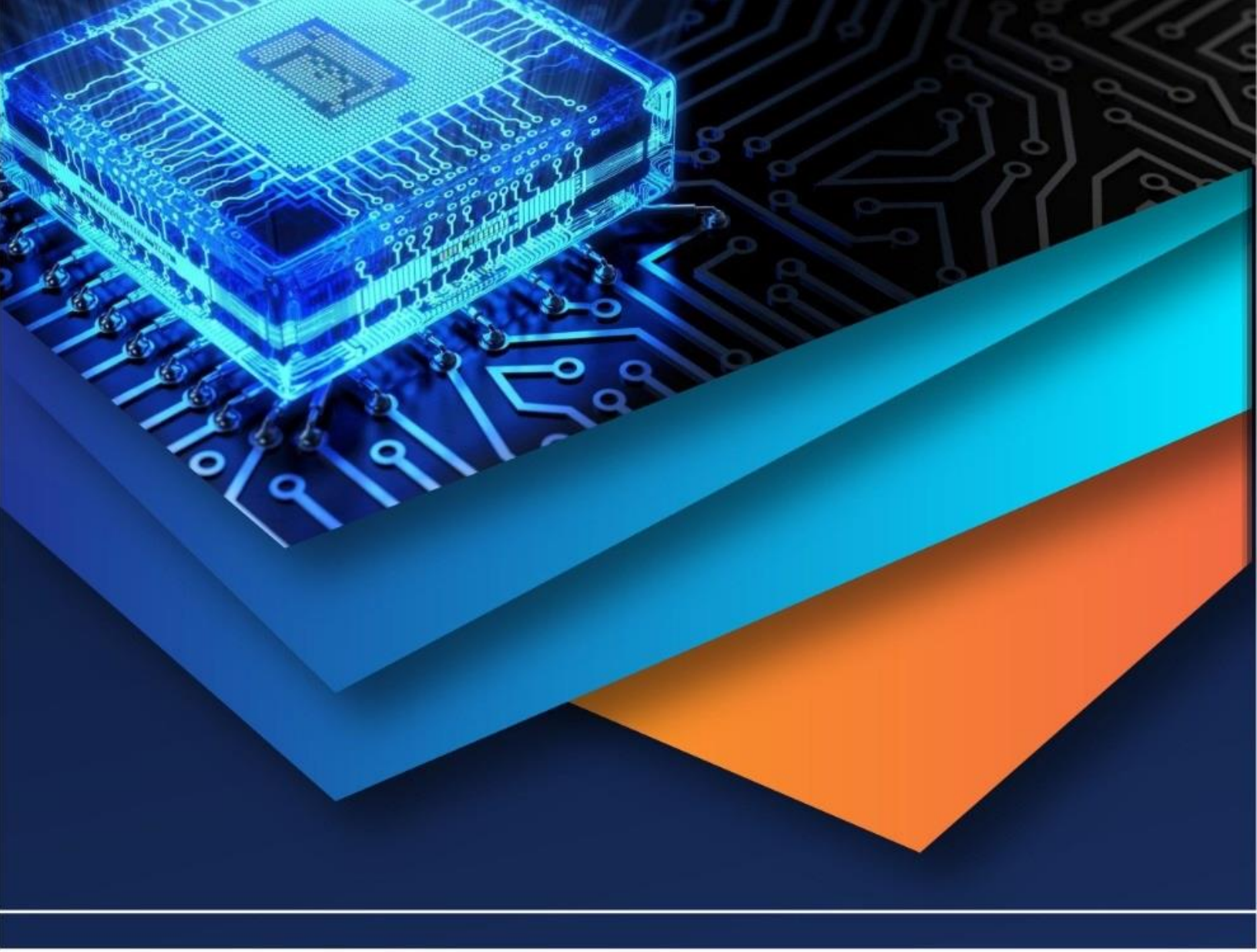
## IX. CONCLUSION

The Decision tree AI estimate is utilised in this work to develop an assumption model for predicting implicit selling costs for any land property. Fresh parameters like air quality and wrongdoing rate were linked to the dataset to aid in predicting expenses even more accurately. This system is fascinating since these elements aren't commonly related with the datasets of other assumption structures. People's perceptions are swayed by these qualities while purchasing a home, so why prohibit them from forecasting housing costs? The Stoner Interface is used in the pre-arranged model, which is built with the Flask Framework. While anticipating the charges at the land expenses, the method provides 89 percent.

## REFERENCES

[1] Lakshmi, B.N., and G.H. Raghunandhan."An abstract overview of data mining."2011 National Conference on Inventions in Emerging Technology. IEEE, 2011.
[2] Manjula, R., etal." Real estate value vaticination using multivariate retrogression models." Accoutrements Science and Engineering Conference Series.Vol. 263. No. 4. 2017.

[3] A. Varma etal., " House Price Prediction Using Machine Learning And Neural Networks," 2018 Second International Conference on Inventive Communication and Computational Technologies, pp. 1936 – 1939, 1936.
[4] Arietta, SeanM., etal."City forensics Using visual rudiments to prognosticatenon-visual megacity attributes."IEEE deals on visualization and computer plates20.12 (2014) 2624-2633.
[5] Yu, H., andJ. Wu." Real estate price vaticination with retrogression and bracket CS 229 Afterlife 2016 Project Final Report 1 – 5." (2016).
[6] Li, Li, and Kai-Hsuan Chu." Vaticination of real estate price variation grounded on profitable parameters."2017 International Conference on Applied System Innovation (ICASI). IEEE, 2017.
[7] Nihar Bhagat, Ankit Mohokar, Shreyash Mane" House Price Soothsaying using Data Mining"International Journal of Computer Applications,
[8] N.N. Ghosalkar andS.N. Dhage, "Real Estate Value Prediction Using Linear Retrogression, "2018 Fourth International Conference on Computing Communication Control and Robotization (ICCUBEA), Pune, India, 2018, pp. 1-5.
[9] Pow, Nissan, Emil Janulewicz, and Liu Dave Liu."Applied Machine Learning Project 4 Vaticination of real estate property prices in Montréal." Course design, Presentation-598, Fall/ 2014, McGill University (2014).
[10] Sampathkumar,V., Santhi,M.H., & Vanjinathan,J. (2015). Vaticinating the land price using statistical and neural network software. Procedia Computer Science, 57, 112-121.
[11] Banerjee, Debanjan, and Suchibrota Dutta."Predicting the casing price direction using machine literacy ways."2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI). IEEE, 2017.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)