ISSN 2278-3091
Volume 9, No.5, September - October 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse295952020.pdf
https://doi.org/10.30534/ijatcse/2020/295952020

Smart Home Value Prediction Using Machine Learning Techniques

Abdul Saleem Javeed[1], Tanveer Ahmed[2], C. Gopala Krishnan[3], Beena G Pillai[4], Madhurya J A[5]
[1]Assistant Professor, Dept. of CSE, GITAM (Deemed to be University), Bengaluru, India, ajaveed@gitam.edu
[2]Assistant Professor, Dept. of CSE, GITAM (Deemed to be University), Bengaluru, India, tahmed@gitam.edu
[3] Associate Professor, Dept. of CSE, GITAM (Deemed to be University), Bengaluru, India, gchandra@gitam.edu
[4]Assistant Professor, Dept. of CSE, GITAM (Deemed to be University), Bengaluru, India, bpillai@gitam.edu
[5]Assistant Professor, Dept. of CSE, GITAM (Deemed to be University), Bengaluru, India, mambuja@gitam.edu

ABSTRACT

The home loan showcase is one among the most market driven and wavering organizations. One of the significant fields in this is applying the AI systems as to how to improve and foresee productivity with high precision. There are three elements influencing a house value which incorporate physical conditions, ideas, and area. The present system includes estimating the house costs with no market cost projections and cost increments. The point in this paper is to anticipate private costs for clients in view of their monetary idea and needs. By separating past business patterns and worth levels, and gaining future ground, costs are anticipated. This test implies anticipating house costs with Random Forest algorithm in Mumbai city. It will assist customers by avoiding paying of cash by going towards an intermediary. The after effect of this exploration demonstrated that random forest gives a slight base mistake of expectation that is 0.349.The experiment results shows that the proposed method is better in prediction and efficiency.

Key words: Housing Prices, Area, Prediction, MSE and RMSE.

1. INTRODUCTION

Studying the worth of immovable property is a crucial aspect in improving the urban structure choices. A delicate stochastic cycle is the land-frame. Choices made by financial experts rely on trends available to produce the most drastic returns so designers are excited by their simple leadership in understanding the possible trends. To accurately measure the cost of immovable property which is a potential trend, massive calculations and use of knowledge results in cost impact and this is needed for analysis, demonstration and determination.

Skyscraping of property costs in metropolitan territories leads to non-deterministic valuation of house costs. In India the equity is depleted as per the dealer's organization. Purchasing a house in India in this manner is a one-sided practice, since this gave minimal uniform type of unveiling the deal cost of the property. By the way broad communications can misuse the material as indicated by their accommodation and benefit.

As a real estate finance specialist, someone needs to consider the hidden aspects of the land display and what influences its prices. As a normal person, a tool that understands these dynamics and impacts of various criteria on property costs is required which can examine in depth the specific business factors and their effect on housing costs. There are several applications in deep learning that could be used to accurately predict appreciations.

We need a model in any case that can forecast potential property predictions with greater precision and less error. For the preparation of model with meaningful memorable data sets metric, we need a common end target. Seeing that less work is being done in India on land proprietary forecasting, we want to build a structure. It will estimate the valuation of the property by considering the different factors that influence the goal value and also to measure statistical performance by considering various error metrics.

2. RELATED WORK

The parts that impact the expense of land must be perceived, and its effect on the costs should likewise be appearing by an audit of authentic information. It is recognized that with this

understanding of the time structure it is not considered appropriate to forecast the development of a clear direct numerical relation. Hence, developing an in-direct framework that fits the information hallmark for evaluating and forecasting future patterns is literally wound up.

Considering that the land segment is increasing in the Metropolitan Region of Mumbai (MMA), inspecting and calculating housing costs using computational show as well as other reasonable techniques is an urgent critical requirement to make the required decisions.

R. Manjula finds any estimation called, sort out dive estimation that fundamentally diminishes the figuring trouble, lessening the measure of features while picking the most critical ones. Affiliations like Zillow.com, magicbricks.com Comprise of a colossal assortment of homes utilizing profound learning and figuring out how to fuse the idea of counts from past analysis.

In the above-mentioned paper designer directed model details were depicted using a lone item computational model using few features as its aspects and model was produced using the input as circumference or compounded and the source normal point bungle for all of the proposed system was calculated (RMS consider).

Nissan Pow had both agreed to seek and sell the expense of lands. For example extra geological highlights were deleted from the Montreal Academic research Repository from the nearest central police command and re-station. For starters, they used approaches including direct relapse, solid vector relapse (SVR), K-Nearest Neighbors (KNN), and Tree / Random Forest Regression. Their finding demonstrated the requesting cost with a mistake of 0.01 utilizing an assortment of KNN and arbitrary timberland gauges furthermore where material the land cost sold was normal with a blunder of 0 023 utilizing the irregular woods regression. Their contention depended on the unobtrusive components of the addresses of guidelines, the survey of land posts and the checking and acknowledgment emerging from the unmistakable gauges right now Hromada discusses the particular proportions of social occasion data from different advancements and land locales and submitting it into various orders. After the affirmation of the subtleties, the model framework will draw up computation plans which will break down the connections between checked factors and delineate the plot see as showed by user requesting A-Z.

AI estimates are used to check true knowledge regarding the emerging Chilean lodging industry. They would presumably take a gander at the judicious usage of the answers to traditional Ordinary Linear Regression for the Neural Network, Random Forest and Help Vector Machine. The illustration for our inquiry provides delineation for new residential units or private property within the guaranteed area of 16,472 worth information. The survey findings indicate that Random Forest worked better than elective ones in showing housing prices. Much of the time, it appears to be assumed that AI tactics will have useful computer game plans for accommodation.

Li Li and Kai-Hsuan studied that the volatility of real estate values has confused habits with some ambiguity and semi-linearity. Writer was using the statistical model-free method of the neural network algorithm. We used back propagation neural system (BPN) and upraised hypothesis feature neural system (RBF) two plans have been used to build Taipei, the dynamical framework of Taiwan for true home meaning variation standards in terms of traveling and parallel number lists. The mean primary interest and root implies two meaning variety lists in squared mistake are picked as the implementation set. Hence, depending on the root of the report, it has been concluded that volatility in housing market trends is not so trust worthy.

Mortgage costs vary continuously, and it requires a system to estimate subsequent house costs. House value forecasting would enable the developer to assess the selling price of a home, which also helps the buyer to schedule the right period to buy a property. And this article aims to forecast house costs in conjunction with the NJOP home in Malang region with regression inquiry and particles colony enhancement PSO. PSO is used to select effect variables, and a relapse model is required to assess the optimal prediction coefficient. The inference from the proposed paper provides the discovery-shown combination relapse and PSO is logical, and the least error is IDR 14.1.

The assessment about its expense is noteworthy as the land costs continue advancing in this manner. At the present time makers tackle two fundamental levels in sum. The essential stage positions a get-together of client described locales to find an ideal region and the resulting stage chooses the most appropriate zone as demonstrated by their necessities and pressure. This follows a customary strategy named clear backslide, which offers an investigation of the got impacts. It assists work with increasing the nature of the association between subordinate variable and other developing self-sufficient variable,

distinguished by and large as name trademark and configuration land. Relapse exhibits away from of the needy quality for example name consistency utilized for the conjecture.

The journalists eventually conducted a comparable analysis of numerous AI-based tests to be reliable Random Forest utilizing incline plunge, K-nearest neighbor backslide, and Random Woodlands backslide to evaluate trends of land confidence. The purpose of the whole article is to discuss the common sense of these AI situations, and to pick the most accomplishable.. Parameters such as outdoor space, quantity of quarters, good way from air terminal / press way / noteworthy, accurate emphasis, closeness to mending offices, retail decisions, number of theaters, field location are used to achieve the argument, as the yield vector is the dedication to the layout and the price of land in the corresponding houses.

The quarterly description in mid 2005-2016 was used to create the model when knowledge compilation and knowledge are collected from sites like 99acres.com, Magicbricks.com, Google.com via Web Scraping. The evaluation is conducted onto the collection with preparation details to concentrate on the various figures using the cumulative objective point error map of the root mean squared fault and the real total loss mea calculation.

Because specific work is undertaken by scholastics utilizing numerous AI technologies, the estimation of land costs is seen as a challenging investigation. The test reveals specific findings from both of the surveys, but the missing factor is that it does not predict potential use of the rooms decided by the customer. The danger issue, despite such findings, is a full addition to a condominium or field. Customers want to find a distributor to that the error, which would increase the system's cost again. That converts and strengthens the present framework.

3. FRAMEWORK PLAN AND ENGINEERING

The plan is isolated in three primary phases: Beginning, Center, Last. Information processing and interpretation define the initial stage. The center stage consists of various sub-stages such as collection of functions, testing of the model of Random Forest and show of SVM, validation of the model and calculation of error. Last Stage requires Final Material Visualization.

Phase 1: Collection of information

Collection of information is the mechanism by which the statistics are measured and collected using a database data collection technique and various procedures. We need to obtain the quantitative data that is ordered and classed. The Indian Government maintains notes and documents on land house value lists on the internet. These documents are made up of verifiable knowledge from earlier years and would be helpful in certain main objective. Sorting of information is needed before some kind of AI exam is conducted. If the validity of the information is in any case an essential requirement there is little need to analyze the results. So we ought to be careful about the information's wellspring and test its validity so far.

Phase 2: Information Cleaning and Loading

Cleaning the information is the strategy for tidying up our informational collection. We may have separate trash esteems in the collected information. Such trash esteems can be erased by testing if the information has any missing qualities or not. We also need our dataset's validity to end. In a given range the values also need to be present. If there are many missing values in a variable we can drop those values. We will need to establish the values before the algorithm **Figure-1** is applied, since each parameter has specific units, so the value is not standardized and the details must be uniform.

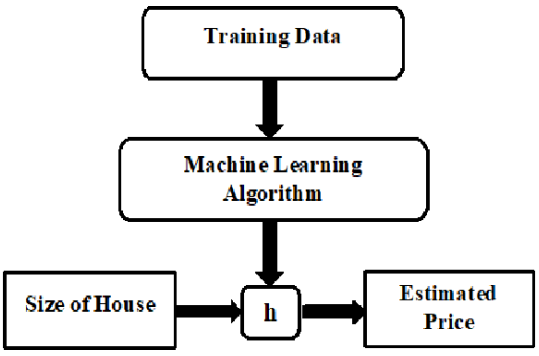

Figure 1: Information Cleaning and Loading

Phase 3: Feature Selection

Additionally the set of capabilities is called vector accuracy. It is a methodology where the collection of factors or elements from an enormous standardized system of specifications takes place in subsets. Structure design methods are used with three purposes: to simplify the method, to reduce the planning period of the layout, to restrict capacity.

Phase 4: Train LR model

As the information is separated through double modules: Learning package and Test set, it will first practice the prototype. There should be one

goal element in the Learning kit. So Random Forest attempts to the gradient in line with the defined parameter with minimal error.

Phase 5: Approval of model

It is the approval technique for deciding if the measurements are performed as per the predefined model. Afterwards, the cohesiveness of model would be as good as would be predicted under the circumstances. When the measurement is completed we will check how well findings match our pattern.

We may even add at least two models to check the norm, or which one is good for certain model. In these analyzes the technique is used as an input yield phase. The acceptance technique analyzes the yields of the gadget being evaluated in comparison to the yields provided from the inquiry, offering the process a comparable data criterion. It preserves the gadget's performance valuations.

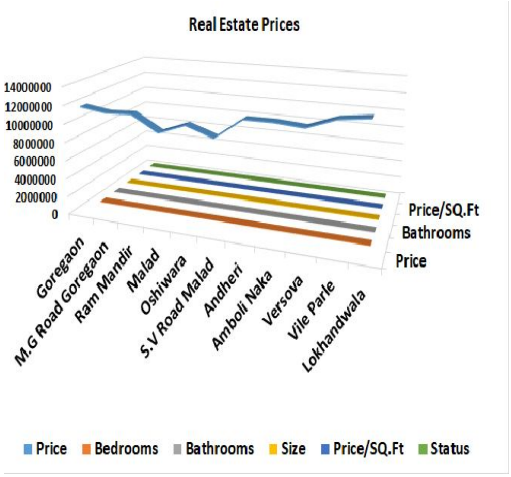

Figure 2: Places v/s real estate prices

4. METHODOLOGY

Anticipating ardent property projections needs a tremendous range of variables, such as region, urban vicinity, range of floors, time period of usability, particular residential units, number of beds, toilets available, secure parking place, lift, layout type, aggregate or capacity, overhang land, building efficiency, cost per unit of capacity **Figure-3**.There are frequently specific parameters that decide the expense of a co-identified house. Using Random Forest System is problematic in this direction the use of multiple ward variables. In prescient review Random Forest is extremely valuable tool.

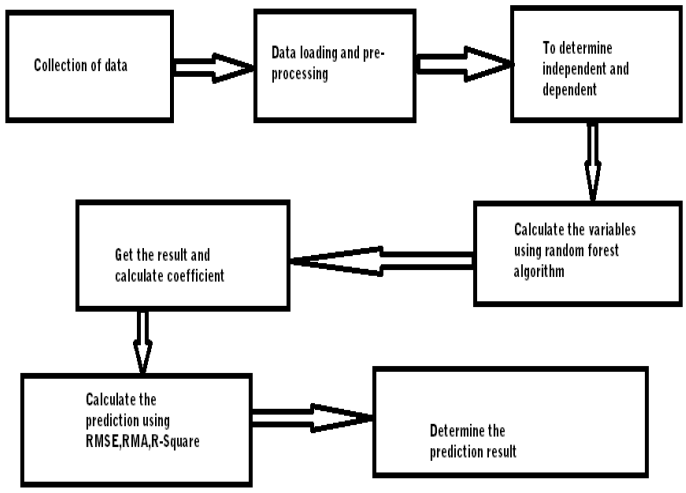

Figure 3: Methodology and properties with their sorts

5. ALGORITHM USED : RANDOM FOREST

The index of property prices Figure .2 and Figure 4 includes attributes such as percent, higher, average, and bottom. The higher segment consists of standard house estimates that are huge at cost, correspondingly usual and lower consisting of traditional center field estimates and low range house estimates. Taking into consideration the end goal of utilizing straight relapse the quarter function is allocated on the velocity vector and estimates of the y axis pace. Direct relapse is performed once for each of quality. The x-hub is self-ruling is the customer's choice quit a drop-down set.

Figure 4: Price Deviations

In Random Forest we conclude a relation between the self-sufficient function vector and the destitute target matrix. Using the free specifications we will predict objective feature. The knowledge element which is

self-sufficient can be a vector with parameters or properties in N. Otherwise, they are labeled regressors. This occurs the correlation between the inferior function and the regressor is obvious. The exacerbation of foreseen confidence and the excitement being observed is known to be botch the subsequent stage is the most commonly viewed method for distinguishing the remaining estimate of space is between the variables better synchronized with the link hand. This procedure builds up the differentiation between watched data (genuine regard) and its vertical division from the planned best fitting line (foreseen regard). It squares every qualification, and incorporates each.

The MSE (Mean Squared Error) is a consistency metric for the evaluator by splitting the RSS and stacking up to the measured information targets. It is trustworthily a non-negative figure. Characteristics are look like zero refers to a littler blunder. RMSE (Root Mean Squared Error) is the square basis of the MSE. The RMSE is a measure of the normal deviation from the figures used in the tests. This is less difficult to consider compared with MSE, which may be a large amount. Dissimilar to different models which can just group the yield, Random Forest will foresee the specific numerical objective worth. In this way Random Forest assumes a solid job in anticipating the value estimation of land.

6. RESULT

```
Coefs:
[[ 1.51258936 -0.66501464 -0.14479887 -0.00513912 -0.00620638]]
Intercept:
[ 16.94255589]
Residue
[ 69.2351298]
Residual sum of squares: 69.24
Variance score: 0.87
(Natural Log transform) Original price 16.91 with the predicted one 16.67
(Real) Original price 22000000.00 with the predicted one 17354924.59
--------------------------------------------------
gross_area
[0 1 2]
saleable_area
[0, 3]
nb_bedrooms
[0, 4]
nb_bathrooms
[0, 5]
(Natural Log transform) Original price 16.91 with the predicted one 16.78
(Real) Original price 22000000.00 with the predicted one 19317070.46
Coefs:
[[ 16.94255589]
 [  0.56970874]
 [ -0.27104379]
 [  0.34907431]
 [  0.04527904]
 [  0.03586186]]
Residual sum of squares: 80.26
Variance R^2: 0.85
```

7. CONCLUSION

In the current real-estate scenarios it has been demonstrated that to test, to store and sort out immense amounts of information based on precondition, different information would be useful so the framework permits ideal utilization of the arbitrary timberland algorithm. This permits the most proficient utilization of such subtleties. Direct apostatize tally satisfies customers by extending their preferred precision and diminishing the risk of putting resources into a home. A huge amount of things that might be added to make the structure all the more satisfying.

REFERENCES

[1] Gang-Zhi Fan, Seow Eng Ong, Hian Chye Koh "Determinants of House Price: A Decision Tree Approach", November 2006**.**

[2] Ali Uzpak "Finding Advanced Features for Housing Price Prediction", August 2017.

[3] Stephen Law, Brooks Paige, Chris Russel "Using Street View and Satellite Images to Estimate House Prices", October 2019.

[4] Robert Thas John **"**Simple Housing Price Prediction Using Neural Networks", May 2018.

[5] Wan Teng Lim, Lipo Wang, Yaoli Wang, Qing Chang "Housing price prediction using neural networks", August 2016.

[6] P. Durganjali, M. Vani Pujitha "House Resale Price Prediction Using Classification Algorithms", March 2019.

[7] Jiao Yang Wu "Housing Price prediction Using Support Vector Regression", May 2017.

[8] Daniel Park, Dimitri Liakhovitski, Gwendolyn Fernandez, Henry Crosby "Predicting Iowa House Prices using Supervised Machine Learning Algorithms" November 2017.

[9] Santhosh Shetty "Predicting Housing Prices — A Random Forest Analysis, March 2018.

[10] Emmanuel Sibanda "Predicting the price of houses in Brooklyn using Python", March 2019.

[11] Shreyas Raghavan **"**Create a model to predict house prices using Python", June 2017.

[12] Abhishek Kumar "Make your own model to predict house prices in Python", Jan 2019.

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

[14] Subham Sarkar "Predicting House prices using Classical Machine Learning and Deep Learning Techniques", September 2019.

[15] Yulia Norenko, Howard Chang, Lavanya Gupta, Silvia Lu "Predicting House Prices with Machine Learning Algorithms", August 2018.

[16] Sara Gaspar "Predicting house prices with linear regression", January 2018.

[17] Hrishikesh Kale "House Price Prediction Using Regression Algorithms", March 2019.

[18] Randerson "Predict Boston House Prices Using Python & Linear Regression", June 2019.