

# HOUSE PRICE PREDICTION USING MACHINE LEARNING

**Animesh Kumar Singh<sup>1st</sup>**

Department of ECE

Sreenidhi Institute of Science and Technology,  
JNTUH, Telangana, India

**C.Lokesh<sup>3rd</sup>**

Department of ECE

Sreenidhi Institute of Science and Technology,  
JNTUH, Telangana, India

**A.V.S Karthik<sup>2nd</sup>**

Department of ECE

Sreenidhi Institute of Science and Technology,  
JNTUH, Telangana, India

**Dr. V Kumara Swamy<sup>4th</sup>**

Department of ECE

Sreenidhi Institute of Science and  
Technology, JNTUH, Telangana, India

## ABSTRACT:

Sharp declines or sustained upward trends in property prices have attracted the attention of many other stakeholders, not just researchers. There are several research papers that use various methods and solvables to solve real estate price change problem. This paper addresses changes in house prices as a classification problem and describes a machine learning approach that uses existing data to forecast the rise or decline of housing values. In the given paper, we employ a variety of feature selection strategies, including the variance effect, different given value, and principal component analysis, as well as Methods for transforming data, including methods for addressing outliers and missing numbers. Accuracy, precision, specificity, and sensitivity are the four metrics used to evaluate the effectiveness of machine learning algorithms. Real estate is not only a basic human need but also represents the wealth and prestige of a person today. Investing in real estate generally seems to be profitable because these properties have not depreciated quickly. Changes in real estate can affect many home investors, bankers, policy makers and more. Investing in real estate seems like an attractive investment opportunity. Therefore predicting the value of real estate is an important economic indicator. Design methods include data preprocessing, creative feature design, and regression models, such as brush regression, gradient boosting, linear regression, and hybrid regression

**Keywords:** Machine learning techniques, Regression analysis, Information value, Python modules, Specificity, Jupyter.

## 1.INTRODUCTION:

The development of civilization created the basis for a constantly increasing demand for housing. Accurately predicting property prices has always fascinated Bankers, purchasers, and sellers. The real estate price prediction conundrum has already been attempted by numerous scholars. As a result of the research activities of various scientists around the world, many theories have appeared. Several of these ideas suggest that a region's topography and culture influence whether housing prices rise or fall, while others emphasize the socioeconomic conditions behind the rise in housing prices. We all know that house prices are within a certain range, so predicting house prices is clearly a regression problem. Generally, to predict home prices, people try to find similar properties in the neighborhood and use the information gathered to predict home prices. All of this suggests that predicting home prices is a new area of regression research that calls for knowledge of machine learning. This inspires me to pursue a career in this area. A real estate appraisal is an important part of the real estate buying process. Traditionally, appraisals are done by appraisers with special training in real estate appraisal. Automated price estimation systems help real estate buyers approximate the prices of real estate present on the market. Such systems are especially useful for first-time home buyers with little or no experience.

### A. Literature Review

Article Title: Virtual Reality in landed property  
 Authors: Luminita Parv, Alexandru Deaky, and Bogdan. The outcomes of the VR RE study are presented in this article. The project's objective is to help real estate buyers and sellers save time and money. using the latest technology. VR

RE is one of Blue mind Software's innovative projects and is currently in progress. This post also explores the history of our internal technical efforts to create real estate friendly presentation tools using 3D and VR (Virtual Reality).

2nd Paper Title: Creating Intelligent Commercial Real Estate Written by Peter Ekman The given paper describes using CNN for detection and assessment patients who are infected. CNN Adapted Models: CNN-based adapted models were developed to cover three image categories (normal cases, viral pneumonia cases, bacterial cases, etc.). An RNN-based architecture called DenseNet169 is employed. We examine a Swedish commercial real estate that created and implemented technology-based selfservice (TBSS) to assist tenants in lowering their energy use in order to evaluate the potential of Smart Commercial Real Estate (CRE).

3. Title of the article. Particle Swarm Optimization (PSO) Based Optimized Support Vector Machine (SVM) for Cryptographer Prediction Nor Azizah Hitama, Amelia Ritahani Ismailb, and Faisal Saeed are the authors. For Predicting the Cryptographies Swarm Optimization (PSO). Forecasting is crucial to the banking industry. An optimised To forecast future cryptocurrency prices, a support vector machine (SVM) based on particle swamp optimisation (PSO) was developed. It is a piece of artificial intelligence (AI) that forecasts future prices based on historical data.

.Article Title: Comparing Ensemble Techniques for Property Valuation

Authors: Prathamesh Kumar, Ishan Madan, Ashutosh Kale This article mentions gradient boosting, random forest, severe gradient boosting, and bagging. We examined four ensemble approaches. Compare Mumbai property valuation efficiencies. Real estate The data for this study was derived from listings on the 99 Acres real estate website. When compared to other Ensemble models, the Extreme slope Boosted model performs the best. This finding validates the ensemble model's utility in calculating real estate prices.

### B. Limitations of existing systems

There are several property classification websites in India listing properties for sale/buy/rent. However, each of these websites has many differences when it comes to apartment prices. Similar properties may be priced differently due to insufficient translucency and precision. A client may feel that the value of a particular listed home is not justified, but there is no way to ensure and verify that the details are correct. It can bring more translucency and trust to the sector. This is especially essential for most Indian buyers because the transaction price is very extreme and any solution to this problem will offer help. Long term, for both our clients and the real estate industry. We suggest that computational modeling and artificially intelligent approaches be used to create algorithms that can assess housing prices depending on certain input data. This individual's commercial use has been that classified sites can utilise it directly., taking some input variables and predicting accurate and fair prices for newly listed properties Predictable.. Do not give your home a false appraisal. This research provides evidence of idea and might be considered an assessment report. There is also a need to enhance the understanding of proper real estate appraisal through accurate appraisals. No price entries and customer error creeps into the system. To knowledge of ours , this work on aggressive real estate pricing in the Indian contexts has never been published in the literature. and contests are held. B. Kaggle's Boston House Price Challenge.

### C. Proposed System

Machine learning refers to the use of data by computer systems to improve their performance and learn without human intervention. This technology is widely used to build predictive algorithms, perform various computational tasks, and develop difficult models. The primary objective of machine learning is to enable computers to learn from data without being explicitly programmed. It involves providing data and using different algorithms to build and train models. Applications of machine learning include facial recognition technology, among others. A Svm Classifier is a prominent method for supervised learning of computer vision (SVM). The Sigmoid activation function is used for categorization as well as regress. problems and aims to create a minimum line of decision boundaries that can divide n-dimensional spaces into classes. SVMs are known for their high accuracy and fast prediction performance compared to other algorithms like Bayesian

algorithms. It also require less space because the judgment step only uses a fraction of training data. However, SVMs do not Although the input chances are not category chances, chance calibrate techniques may be employed to turn the data to classifier. For binary classification, probabilities can be adjusted using Platt scaling, which involves regression logistic of SVM values and additional cross-validation on training data. In real-world applications, SVM machine learning technology can be used to To deliver the most realistic cost to customers, we estimate home values depending on many variables.

D. Need and motivation

After living in India for many years, I noticed that housing and rent prices continued to increase. However, it was surprising to learn that in the fourth quarter of 2016, property prices in Bombay experienced the largest drop in four years. This decline resulted in a 6.3% decrease in average resale prices for condos and co-ops, which was the first decline since the first quarter of 2017. The decrease was partly due to political uncertainty, both domestically and internationally, and the 2014 election. To ensure transparency with customers and enable easy comparisons, a model can be used to predict house prices. If a customer finds a higher price for a house on a particular website than the price predicted by the model, they may choose to reject the house.

2.DATASET:

Here we have web scrapped the Data from

<https://www.kaggle.com/datasets/muhammedabdulazeem/house-rent-prediction-for-hyderabad-website>

All Hyderabad Homes are included in our database. Dataset looks as follows as shown in fig 2.1 and fig 2.2.

[2]:	area_type	availability	Location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Nizampet	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Hitech City	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Manikonda	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Alwal	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kukatpally	2 BHK	NaN	1200	2.0	1.0	51.00

Fig 2.1

]:	Location	size	total_sqft	bath	price	bhk
30	Kokapet	4 BHK	2100 - 2850	4.0	186.000	4
122	Nizampet	4 BHK	3067 - 8156	4.0	477.000	4
137	Manikonda	2 BHK	1042 - 1105	2.0	54.005	2
165	Gachibowli	2 BHK	1145 - 1340	2.0	43.490	2
188	Narsingi	2 BHK	1015 - 1540	2.0	56.800	2
410	Miyapur	1 BHK	34.46Sq. Meter	1.0	18.500	1
549	Kukatpally	2 BHK	1195 - 1440	2.0	63.770	2
648	Gandipet	9 Bedroom	4125Perch	9.0	265.000	9
661	West Marredpally	2 BHK	1120 - 1145	2.0	48.130	2
672	Alwal	4 Bedroom	3090 - 5002	4.0	445.000	4

Fig 2.2

A. Data Exploration

To assess the precision on new, unseen data, it is usual practise in machine learning to divide the information into two separate test sets. The model is trained using the training set, and its performance is evaluated using the testing set. The 9:1 split means that 90% of the data is used for training and 10% is used for testing. This split ratio can be adjusted based on the dimension of the data set and the complexity of the model. It is important to ensure that the split is representative of the overall data set and that there is no bias in the training or testing sets. Cross-validation techniques can also be used to further evaluate the model's performance and prevent overfitting. Data exploration is a critical step in the data analysis process, involving the thorough investigation and understanding of a dataset's main characteristics, including its dimension, accuracy, initial patterns, and other attributes. The process helps in identifying patterns, relationships, and trends within the data that can be used to drive valuable insights and inform business decisions.

B. Data Visualization

Present the data representation of knowledge and data. Data visualisation tools provide an incredibly simple way of discovering trends as shown in Fig 2.3, abnormalities, and data patterns by using visuals including such charting, diagrams, and map. With the enormous quantity of data accessible inside the Big Data era, data visualisation methods and instruments are crucial for analysing massive volumes of information and reaching information decisions.

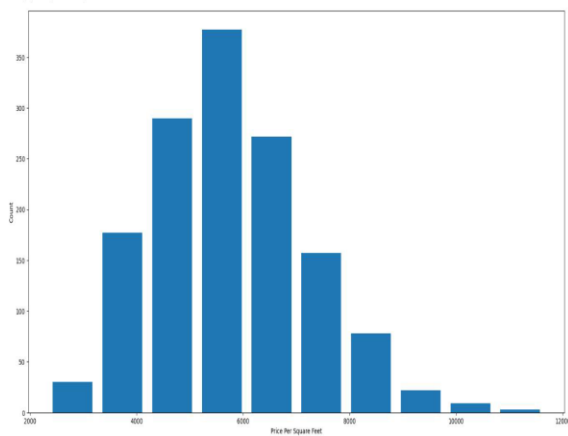


Fig 2.3

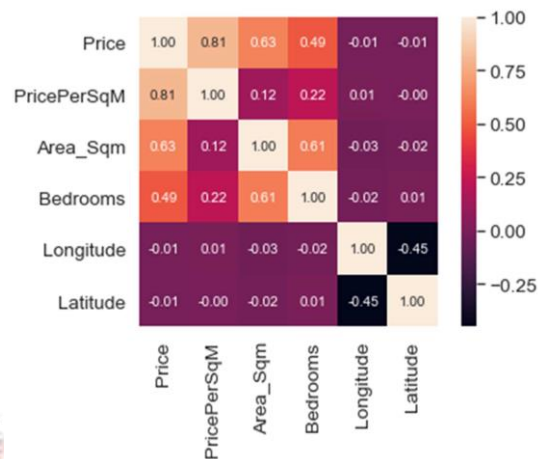


Fig 2.5

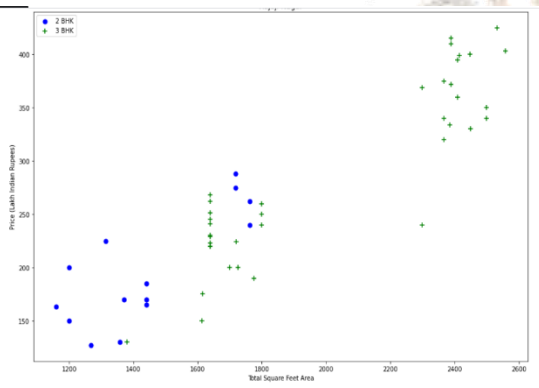


Fig 2.4

C. Data Selection

Data selection is a crucial step in the research process that involves identifying the appropriate data type and source, as well as the most suitable instruments for data collection. It is important to note that data selection is distinct from selective statistics (leaving data that doesn't meet a research question) to real - time data selecting (using gathered data to track actions or secondary analyses). The quality and integrity of data can be affected by the process of data selection. The main goal of data selection is to determine the most suitable data type, source, and instrument(s) that enable investigators to effectively address research questions. This determination is typically influenced by rigour considerations, previous literature, and accessibility to relevant data sources. The selection of appropriate data for a research project is crucial as it can affect the integrity of data. The primary purpose of data choosing is to identify the best file format, supplier, and objects addressing research inquiries.

3.MODELS USED:

Regression Model

- Sequential Vector machines computational model learning - based..
- It did perform an approach to predict. Regression is used to predict a goal predicted values based on subjective parameters
- It primarily serves to determine how factors and predicting the future interact.. Fig 3.1 shows Real Vs Predicted graph.

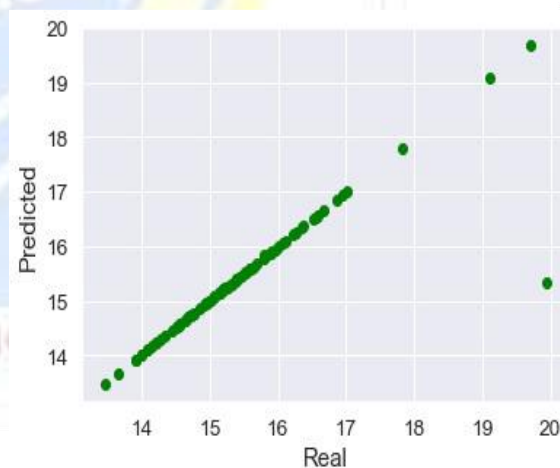


Fig 3.1

4.WORKING:

A software-based online platform called Jupyter, where all the Python modules and routines are developed, is used to test the functionality of this model.

1. Go to Jupyter notebook where the project is built, the windows would look like this as shown in fig 4.1.

```

[50]: from sklearn.linear_model import LinearRegression
lr_cif = LinearRegression()
lr_cif.fit(X_train,y_train)
lr_cif.score(X_test,y_test)

[51]: 0.8642643016357112

[52]: def predict_price(location,sqft,bath,bkk):
loc_index = np.where(X.columns==location)[0][0]
x = np.zeros(len(X.columns))
x[0] = sqft
x[1] = bath
x[2] = bkk
if loc_index != 0:
if loc_index != 1:
x[loc_index] = 1
return lr_cif.predict(x)[0]

[53]: predict_price('Kukatpally',1000, 2, 2)

/lib/python3.10/site-packages/sklearn/base.py:409: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
warnings.warn()

[54]: 61.30910929056554
    
```

Fig 4.1

Then Type Predict\_price('Place name', sqft, bathroom, bedroom)

Then the compiler would give the output as result in lakhs and display it in next box, the result will be accurate upto 10 digits which may help users to plan their budget accordingly. In the above box 61.3091XXX is the price of the given plot in lakhs.

Similarly other places price can also found and price of the same can be predicted.

*A. Results and Discussion*

**Best Suited Model**

Linear regression has shown the best performance on this dataset and can be used for deployment purposes. The coefficients of a linear model comprising one or more different factors that are most accurate at predicting the values of the predictor variable are estimated using this analysis approach. A linear model or area that minimises the difference between the expected and observed output is fit using regression analysis. The sum of squares approach is used by my straightforward regression analysis calculator to identify the best fit line for a set of data sets. then

extrapolate V's value to the corresponding value for U (the predictor variables).

It sounds like we have a tool that can perform simple linear regression analysis using the least squares method. Determining the parameters of a linear model that can most accurately predict the worth of a dependent variable depending on one or more relationship between variables is the goal of this type of investigation. By matching a straight line or range, the aim is to reduce the discrepancy between the anticipated values and the actual output values. With this tool, you can Using the different variables, determine the amount that is depending on variables (V) (U). In table 4.2 different places has been compared with actual vs predicted prices with the help of this model.

S/N	Actual Price (in Lakhs)	Predicted Price (in Lakhs)	Difference
1	34.423	33.343	1.080
2	23.441	21.453	1.988
3	47.532	47.244	0.288
4	19.221	20.242	-1.021
5	54.321	52.352	2.001
6	32.332	32.454	-0.122
7	43.4324	43.2422	0.1902
8	53.322	53.111	0.211
9	24.424	25.442	-1.018
10	21.342	21.543	-0.201
11	43.353	45.324	-1.971
12	58.644	59.455	-0.811
13	62.342	61.352	0.99
14	34.324	35.232	-0.908
15	54.435	53.353	1.082

Table 4.2

**5.CONCLUSION:**

The cost of real estate is increasing annually, and there is a need for a system to anticipate future real estate costs. Home Price Forecast can assist property owners, evaluators, and policymakers in determining home evaluations and reasonable selling prices. This can aid potential buyers in determining the optimal time to purchase a home. The three primary elements that affect home prices are physical condition, style, and Loacat, but the specific variables that influence home prices vary. Therefore, a comprehensive forecasting model should account for the particular variables that impact home prices in the relevant area. The effectiveness of random forest machine learning methods in predicting house prices was further demonstrated in this research, using variables from the Boston Housing Dataset. The model achieved a predicted difference of ±5, indicating that it can be utilized to forecast house prices. Additionally, other

machine learning models, notably deep learning models, can be employed to anticipate house prices.

## 6.COMPARISONS:

There are various websites which predict the price of houses and also gives the details for the same. These are the few websites which are available Online to predict the prices of the houses in the given area. The given websites are new and old generations property buying and renting websites where millions of users use it to buy or rent property on daily basis.

1. *99acres.com*
2. *Magic Bricks*
3. *Housing.com*
4. *Nobroker.in*

Different websites have different outlook and working systems, which comprises their user interface, data set, prices of the houses, number of house, user filtered allocations etc. In todays era data protection from theft is most important and required, Our project doesn't collect any data from the users and directly searches for the house before entering any personal details of the person. Predicting house prices is critical to increasing real estate efficiency. As in the olden days, home prices were determined by calculating local market prices. Therefore, house price forecasting models are very important to fill information gaps and improve real estate efficiency. This model can be used to better predict prices. The Model used in our project is regression Analysis which is one of the most efficient analysis to predict house prices and also used in Deep learning and machine learning.

*A. Issues resolved by our project are:-*

- 1.No Registrations are required before searching for any house or renting an apartment.
- 2.Easy to use where any user of any age can use.
- 3.Datasets use in here are updated and gives up to date results.
- 4.No unnecessary menus and banners are displayed
- 5.No ads Displayed since more than 50% users run away by just seeing ads on the pages.
- 6.Simple layout.

## 7.REFERENCES:

- [1] David Donoho. "50 Years of Data Science". In: Journal of Computational and Graphical Statistics 26.4 (2017), url: <https://doi.org/10.1080/10618600.2017.384734>.IEEE,pp.1.6 pp. 745–766.
- [2] Sameerchand Pudaruth. "Predicting the Price of Used Cars using Machine Learning Techniques". In: International Journal of Information & Computation Technology 4 (Jan. 2014) IEEE 2019 pp. 1.8
- [3] Alejandro Baldominos et al. "Identifying Real Estate Opportunities Using Machine Learning". In: MDPI Applied Sciences .IEEE (Nov. 2018)
- [4] Johan Oxenstierna. Predicting house prices using Ensemble Learning with Cluster IEEE Aggregations. 2017.
- [5] Wikipedia. k-nearest neighbors algorithm. [https://en.wikipedia.org/wiki/Knearest\\_neighbors\\_algorithm](https://en.wikipedia.org/wiki/Knearest_neighbors_algorithm).,IEEE, 2018 pp 1.6