International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Property Price Optimization Using Machine Learning Techniques

Devnandan Bihari[1], Nikhil Kumar Yadav[2], Lavkush Gupta[3]
Department of Computer Science and Engineering
IIMT College of Engineering, Greater Noida

devyadav8738@gmail.com , nnikhilyadav147@gmail.com , lavkushgupta96092@gmail.com

Abstract –

As we in this era property price optimization Is a censorious aspect of the real estate industry, buyers, sellers, etc. Traditional models were often failing to account for the complex interactions between various datasets. This study explores the application of machine learning techniques, specifically Random Forest and Linear Regression. The dataset consist of real estate listings with different parameters like square footage, No. of bedrooms , pricing, furnishing status and so on. We have used different ML models such as Linear Regression is a fundamental tool for real estate valuation, providing insights into property trends whereas Random Forest is an learning method that builds multiple decision trees and help to improve accuracy and to handle complex relationships within variables. The models were assessed by using performance metrices such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The result shows that Random Forest model surpassed Linear Regression Model by handling all the errors effectively for future basis. The main motive of this paper is to conclude the models in such a way that it can be easily used by all the writers who are working on this topic. We have summarize all the research paper in one paper.
Keywords- Linear Regression, Random Forest, Price, Furnishing Status, Dataset. .

1. INTRODUCTION

As we know that population Of India is increasing day by day and the demand of property is also increasing day by day. For this instance we are required to have such a model which can help stake holders, buyers and sellers to accurately predict the price of the property [1] . For predicting the price of the property there are several types of Machine Learning Models which can help us in predicting the correct price of the property. Some of the algorithms which can be used for predicting the estimated cost of the property are Linear Regression , Random Forest , XG Boost, and so on. Machine Learning (ML) is a branch of Artificial Intelligence (AI) that helps computers to learn patterns from data and make predictions Without explicit programming. ML algorithms Uses parameters such as No. of Bedrooms, Price, Furnishing Status and so on. Algorithms uses these parameters for estimating the price of the property.

LITERATURE REVIEW-

Over the last years , we have seen the rising of the property prices. So, to overcome the

Problem of property price a lot of machine some of the ways in which ML algorithms can

Be used in property price optimization are- Linear Regression- **Linear Regression** is a

Vital machine learning algorithm which can Be used to model the connection between a
Dependent variable (i.e. property price) and One independent variable (i.e. No. of Bedrooms) . This algorithms always predict Outcomes using Straight Line.
XG Boost- It is a strong machine learning Algorithm which is based on gradient Boosting . It builds multiple trees in a Sequence which makes it easier to predict The price of the property. It is also used to Reduce the problem of overfitting.
Decision Tree- It splits the data into branches Based on feature values which forms a tree-like structure to make accurate Predictions .

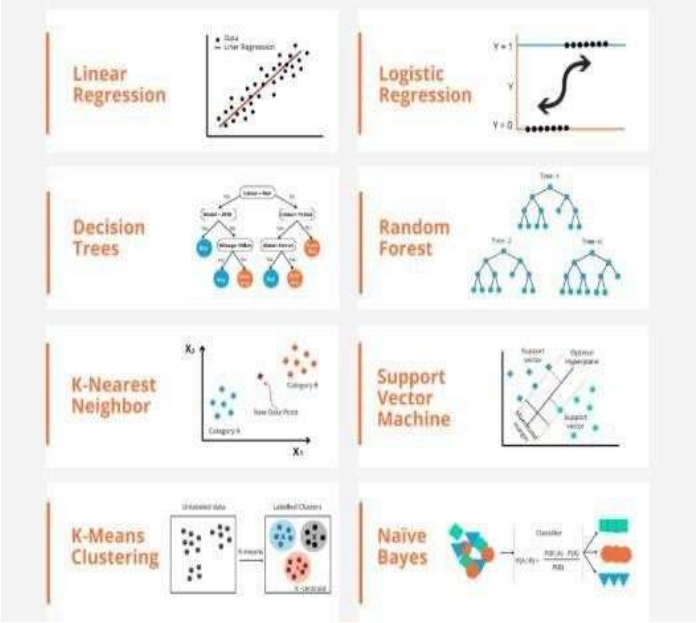
Machine Learing is very important aspect in Estimating the property price using different Algorithm. These algorithms is not only used In estimating price but can be used in reducing The problem of overfitting. It can save time as Well as reduce human power. By offering the parameters like price, square foots, No. of Bathrooms and other parameters we can Easily predict the house price.

We have seen learning models were introduced in the market to help the estate industry, buyers and sellers to take the Correct decision. Traditional Valuation model , such as Hedonic Pricing Model (HPM) and Multiple Linear Regression (MLR), have

been usually used ,but they often fail to capture the complex relationships between features and market dynamics. By using different datasets and different machine learning algorithms we can achieve different results but they are not very accurate in nature . By reviewing the past year research papers we came to the Following Conclusion that-

As we can see in Aldrin's[2] paper they have concluded that how different algorithms can be used for estimating the price of the property .They have predicted the price of the property using different machine learning algorithms such as linear regression, supervised learning and so on.

In Rohit Batchala's paper [3] Focus on urbanization and accessibility to Facilities, this paper uses cutting-edge machine Learning methods including Support Vector Regression (SVR) and Random Forests to forecast House values. We have seen that these Regression techniques effectively identify relationships among features , whereas more Sophisticated methods are adept at managing Intricate patterns. The processes of preprocessing And feature engineering play a crucial role in Boosting model performance. The integration algorithms into ensemble methods Guarantees reliable predictions and flexibility. Utilizing extensive datasets alongside intuitive interfaces enhances Practical insights for real estate valuation and facilities informed decision making among various stakeholders.



The paper by Xiaojie's paper [4] we explore a price optimization challenge faced by brick-and- mortar retailers, taking into account two primary constraints: a limit on the number of price adjustments and a minimum threshold for price changes , thereby ensuring that pricing strategies remain practical. Although the feasible region exhibits non-convex characteristics, the research illustrates that efficient computation of Euclidean projection onto this region is achievable, Which facilitates the creation of a gradient projection Algorithm (GPA) to address the issue. The theoretical Foundations of the algorithm are substantiated , and Computational tests validate its effectiveness. Prospective Research avenues include enhancing the model with Further business constraints, examining decomposition Methods, analyzing mathematical programs with Cardinality constraints, and evaluating alternative Demand functions, such as those derived from discrete Choice models.

In Elias Eze's paper [5] the analysis conducted on Predicting house prices through machine learning Techniques applies to the Boston dataset has provided Significant insights regarding the efficacy, advantages And limitations of various models. The Random Forest (RF) model emerged as the top performing model when Evaluated using the Root Mean Square Error (RMSE) Criterion. The results of this study hold potential Applications in real state , inform policy decisions, and Contribute to ongoing research in machine learning For housing price prediction.

In Dr. Anubhav's Paper [6] they empower middle-class Clients to acquire homes and other properties at their Fair market value, thereby protecting them from Unscrupulous brokers. Furthermore, by proving precise Valuations, this model will enable large organizations to Determine pricing effectively, resulting in significant time And cost savings. Our objective is to ensure that real estate Prices reflect true market conditions, which is fundamental to The industry. The system was designed to predict price by Learning from the raw data it receives. After reviewing various Research papers, blogs, and articles, we selected several Algorithms suitable for both datasets used in the model. Extensive testing and training revealed that the XG Boost Algorithm outperformed its competitors. The system with data and effectively predict the prices of numerous properties with diverse characteristics . The application is both efficient and time-saving..

METHODOLOGY

For property price optimization based on Machine Learning (ML) we are using all the well known algorithms used in Machine Learning. From time to time we have seen different approaches for training our dataset. Whether it is different dataset or different algorithm. In this project also we have use different algorithms for the same dataset. Some of the algorithms which are used in this project are-

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(yi - yi^{\wedge})^2$$

RANDOM FOREST – Random Forest proves to be exceptionally efficient in optimizing property

prices by effectively capturing intricate relationships among various features including location, size, amenities. It is important because- - Accommodates Non- Linearity
Mitigates Overfitting
Analyzes Feature Importance
Manages Missing Data

It is calculated as –

$$y = \frac{1}{n}\sum_{i=1}^{n} Ti(x)$$

where-

y = Predicted Property Price

Ti(x) = Prediction from the i_{th} decision tree

n = Total Number of Trees

LINEAR REGRESSION

Linear Regression serves as a straightforward yet powerful model for price property prices. It is based on the premise of a Linear correlation between independent variables (such as square footage, no. of bedrooms and location) It is important because-

- effectively captures intricate relationships among property characteristics
_ It handles non-colinearity

- Hyperparameter Tuning
Reduces Overfitting It is calculated as-

Where- m = No.Observations

yi = Actual Price
yi ^ = Predicted Price

XG Boost

XG Boost is a sophisticated machine Learning algorithm that utilizes gradient boosting Techniques. It is particularly effective in predicting Property prices, owning to its proficiency in Managing missing data, outliers, and intricate Relationships within datasets.

It is important Because-

- It effectively handles Non- Linearity
- It enhances Boosting Mechanism
- It performs efficiently with extensive Datasets.

It is calculated as-

$$\mathcal{L}(\Theta) = \sum_{i=1}^{n} l(yi - yi^{\wedge}) + \sum_{k=1}^{K} \Omega(f_k)$$

Where-

L(yi, yi^) = Loss Function
Ω(fk) = Regularization term
N = No. of training examples
K = Number of Trees

DECISION TREE

Decision Trees are extensively utilized in predicting property prices because they effectively capture non-linear relationships and interactions among features.

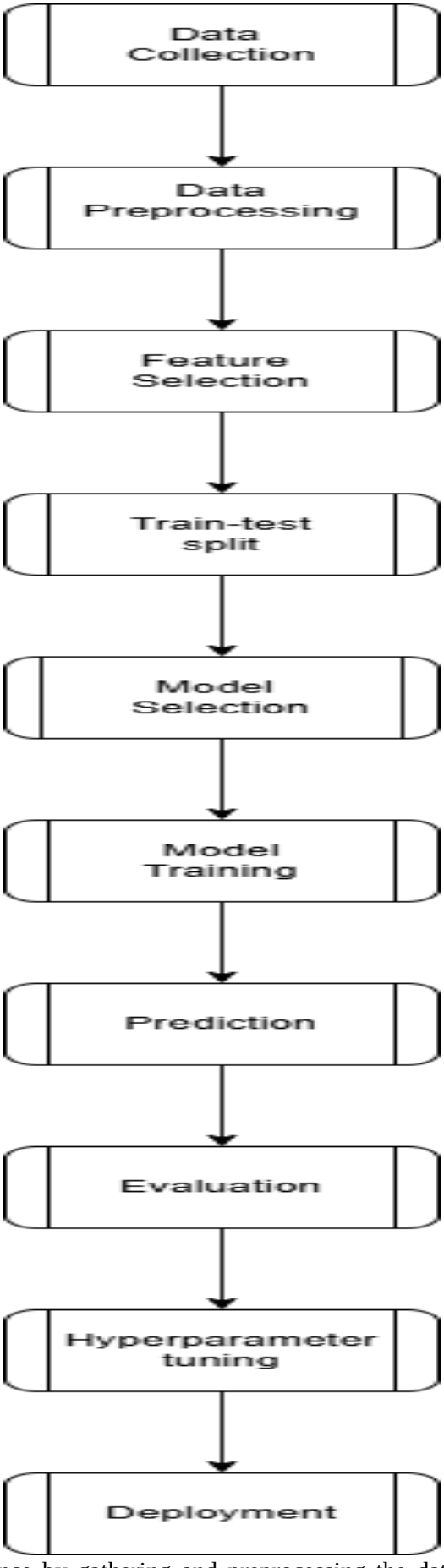
It is calculated as-

$$MSE_{split} = \frac{N_L}{N} MSE_L + \frac{N_R}{N} MSE_R$$

where-
N_L, N_R *= Number of samples in left and right child nodes.*
MSE_R, MSE_L *= MSE for left and right Child nodes.*

Procedures for Model-

To initiate the property price prediction model , we

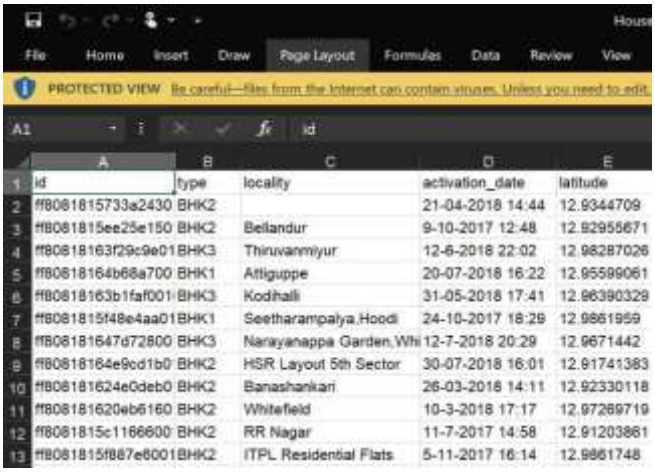


commence by gathering and preprocessing the dataset, ensuring it is well-organized and devoid of any missing values. We identify essential features, Including square footage, number of Including square footage, number of These are the steps which should be follow to predict the accuracy of the model.

Random Forest , or Gradient Boosting is selected And trained on the model's performance is evaluated Using metrices such as Mean Absolute Error (MAE) And Mean Squared Error (MSE), for training the data

FLOWCHART-

1. **DATASET-** choose the appropriate dataset is essential for precise property price forecasting. The data-set ought to encompass vital attribute Such as location, type of property ,etc It is imperative that the dataset is clean, well- Organized , and devoid of any missing or Inconsistent values. Furthermore, incorporating Historical price data and market trends can Significantly enhance the accuracy of the Model. A high quality dataset facilitates Improved feature selection , minimizes biases, And bolsters the dependability of the Predictive model



	id	type	locality	activation_date	latitude
2	ff8081815733a2430	BHK2		21-04-2018 14:44	12.9344709
3	ff8081815ee25e150	BHK2	Bellandur	9-10-2017 12:48	12.92955671
4	ff8081816329c9e01	BHK3	Thiruvanmiyur	12-6-2018 22:02	12.98287026
5	ff8081816 4b68a700	BHK1	Attiguppe	20-07-2018 16:22	12.95599061
6	ff8081816 3b1faf001	BHK3	Kodihalli	31-05-2018 17:41	12.96390329
7	ff8081815f48e4aa01	BHK1	Seetharampalya,Hoodi	24-10-2017 18:29	12.9861959
8	ff8081816 47d72800	BHK3	Narayanappa Garden,Whi	12-7-2018 20:29	12.9671442
9	ff8081816 4e9cd1b0	BHK2	HSR Layout 5th Sector	30-07-2018 16:01	12.91741383
10	ff8081816 24e0deb0	BHK2	Banashankari	26-03-2018 14:11	12.92330118
11	ff8081816 20eb6160	BHK2	Whitefield	10-3-2018 17:17	12.97269719
12	ff8081815c1166600	BHK2	RR Nagar	11-7-2017 14:58	12.91203861
13	ff8081815f887e6001	BHK2	ITPL Residential Flats	5-11-2017 16:14	12.9861748

2.PREPROCESSING – Data preprocessing is a Critical step in preparing the dataset for

Modelling, ensuring that it is both clean and suitable for analysis. This process includes Addressing missing values, eliminating duplicate Entries, and rectifying any inconsistencies present In the data. Categorical variables, such as Location , are transformed into a suitable format, While numerical values may be scaled as necessary. Outliers are detected and managed to Mitigate any potential bias in the model .

3.TRAIN-TEST SPLIT- The division of the dataset Into training and test sets is an essential phase in

The development of a property price prediction Model. This process separates the data into two Segments : the TRAINING SET which is utilized for Model's training, and the TEST SET , which serves to Assess the modal's performance. Commonly, a Split ratio of 80-20 or 70-30 is employed to facilitate Effective learning while retaining sufficient data for Validation purposes. This approach aids in mitigating Overfitting and offers an impartial evaluation of the Modal's ability to generalize to new, previously Unobserved property data.

4. MODEL SELECTION – Model selection plays a vital Role in predicting property prices, as it significantly Influences the precision and effectiveness of the forecasts. The process of selecting an appropriate modal requires a Thorough analysis of the dataset, an understanding of The relationships among features, Of various algorithms based on their performance Indicators. Traditional models, such as Linear Regression, and a comparison Random Forest , Gradient Boosting ,and XGBoost are Adept at capturing intricate patterns and interactions. The modal selection process involves assessing Different modals using metrices like Mean Absolute Error (MAE) and Mean Squared Error (MSE) to Determine the most effective strategy for enhancing Property price predictions Resilience.

5.PREDICTION – In prediction we need to ensure that the accuracy predicted by the model . Key Features include the location, square foot, no. of bedrooms, etc. After preprocessing it should be ensured that the data is cleaned thoroughly to get the required analysis. All the required analysis is done by using different ML algorithm such as Decision Tree, Random Forest , etc. We train the models in such a way that can be evaluated easily.

6. EVALUATION -The assessment of a property price prediction model is essential for verifying its accuracy and Dependability. This process entails evaluating the models ability to generalize to new data by contrasting predicted prices with actual figures. Standard metrics for evaluating regression models include MAE (Mean Absolute Error), MSE (Mean Squared Error), and RMSE (Root Mean Squared Error). A reduced error value signifies a more effective model. Furthermore, methodologies such as cross-validation and feature importance analysis contribute to enhancing the model's precision and Prices. The process entails identifying the most Effective combination of parameters to boost Model accuracy, mitigate overfitting , and improve Generalization capabilities. Approaches such as Grid Search, Random Search ,and Bayesian Optimization are employed to systematically Determine the optimal values for parameters, Including learning rate, tree depth, and the number Of estimators.By meticulously adjusting these Parameters, one can markedly improve the model's Predictive performance , leading to more dependable Estimates of property prices.

7.DEPLOYMENT – Following the training and Assessment of the property price prediction the Subsequent phase is implementation , which Allows for real-time accessibility. This phase Entails hosting the model on a cloud server Or web application, permitting users to enter Property information and obtain immediate price Forecasts. The procedure encompasses model

Serialization , integrating it with a backend Framework (such as Flask of FastAPI), and Deploying it on cloud services (like AWS, Google Cloud, or Heroku). Adequate Optimization guarantees scalability, Efficiency, and a smooth user experience For real-time property price evaluation

RESULT-

The Performance of the trained Linear Regression and Random Forest models was assesses utilizing a test Dataset. Their effectiveness was compared through the metrics of Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 Score . A summary of the results is presented in the Table below.

MODEL	ACCURACY (%)	R^2 SCORE	MAE	MSE	RMSE
Linear Regression	74.06 %	0.7406	3143.4266	18661108.3623	4319.8505
Random Forest	80.81 %	0.8081	2624.1198	13806605.1534	3715.7240

Analysis Of Results-

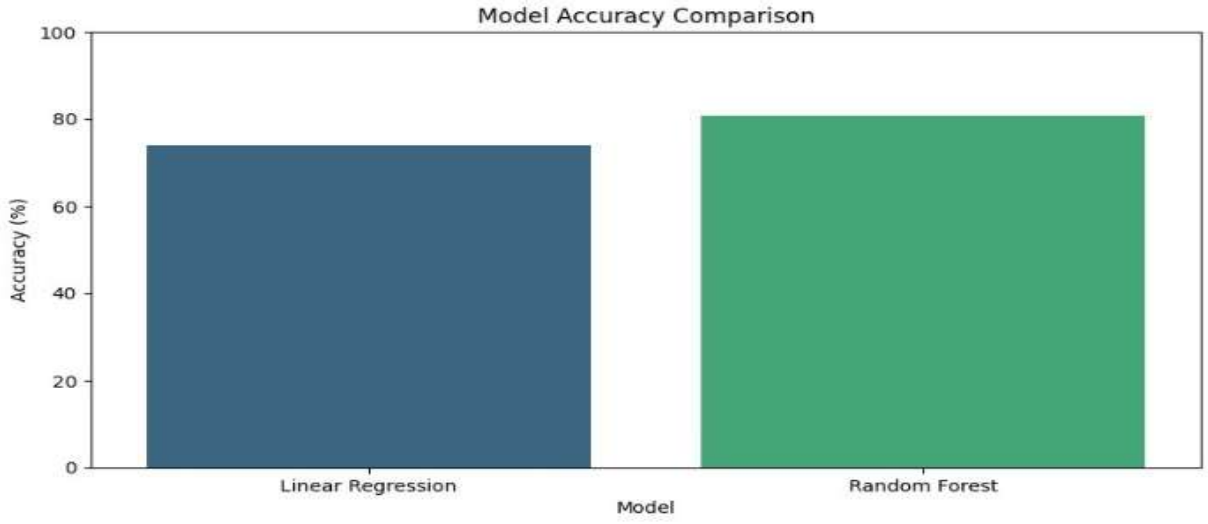
Accuracy:

- Random Forest achieved a accuracy 80.81% , significantly outperforming Linear Regression of 74.06 %.
- The ensemble learning approach in Random Forest provided more precise predictions due to its Ability to handle non-linear relationships in the data.

Error Metrics:

- MAE (Mean Absolute Error): Random Forest (2624.11) is lower than Linear Regression (3143.42), indicating Better precision.
- MSE (Mean Squared Error) & RMSE (Root Mean Squared Error): Random Forest had lower values , proving it generalizes better to unseen data.

R^2 Score: Random Forest (0.808) is closer to 1 compared to Linear Regression (0.740), showing a better Fit to the data.



Graphical Representation of Accuracy Comparison-

CONCLUSION

This research demonstrates that machine learning is effective in accurately forecasting property prices. While Linear Regression offers a fundamental insight into the impact of various factors on pricing, it encounters difficulties when addressing intricate patterns within the data. Conversely, Random Forest significantly outperforms by effectively identifying complex relationships and minimizing errors.

Random Forest is deemed more dependable as it simultaneously accounts for numerous variables and mitigates the risk of overfitting, rendering it a superior option for predicting real estate prices. Additionally, appropriate data preparation techniques, including scaling and encoding , enhance the efficacy of these models

FUTURE SCOPE-

Future investigations may delve into deep learning methodologies, including neural networks, to enhance the precision of predictions. The incorporation of real-time market data, geographic information systems (GIS), And economic indicators has the potential to significantly improve property price forecasting systems.

REFERENCES-

Syafiqah Jamil, T. M. (2020). Machine Learning Price Prediction on Green Building Prices. Cornell Library University (p. 6). Malaysia: IEEE.

[1] Aldrin Fernandes, A. Q. (2021). House Price Prediction Using Machine Learning. International Journal of Advanced Research, Ideas and Innovations in Technology (p. 5). Mumbai: IJARIT.

Batchala, R. (2024). House Price Predictions Using Machine Learning. International Journal of Research Publication and Reviews, Vol 5, no 11, pp 7961-7965 November 2024 (p. 5). Andhra Pradesh: IJRPR.

Wang, X., Huang, H.-C., Han, L., & Lim, A. (2021). Price Optimization with Practical Constraints. arXiv:2104.09597

Elias Eze, S. S. (2023). A Comparative Study for Predicting House Price Based on Machine Learning. 2023 4th International Conference on Data Analytics for Business and Industry (ICDABI) (p. 7). London: IEEE.

Dr. Anubhav Sony, A. (2023). A Strategic Alternative to Predicting Property Investment Value Through Random Forest and Gradial Enhancing. 2023 International Conference on Artificial Intelligence and Smart Communication (AISC) (p. 4). Mathura: IEEE.