Machine Learning for Rental Price Prediction: Regression Techniques and Random Forest Model

Raju[1], Arham Neyaz[2], Aleem Ahmed[3], Annu Singh[4]

[1,2,3,4]Niet, Greater Noida, Uttar Pradesh
[1] raju@niet.co.in

Abstract: In the rapidly evolving real estate market, accurate and data-driven methods for predicting house rental prices are essential. This research paper explores the use of machine learning to forecast rental prices with precision. Traditional approaches to determining rental prices often rely on local market knowledge and intuition, but they struggle to account for the complex interplay of factors influencing rental rates. Machine learning, employing regression analysis and historical rental data, offers a revolutionary paradigm shift in rental price prediction. This paper investigates the impact of variables such as house age, furnished status, flat area, room size, and the number of rooms on rental prices, demonstrating how machine learning models can uncover hidden correlations and provide accurate rental price estimates. The motivation behind this research stems from the lack of highly accurate existing models, the need to mitigate overpricing, adapt to market dynamics, optimize investment returns, facilitate data-driven decision-making, promote market transparency, and develop a user-friendly interface. The primary objective is to develop a robust and accurate house rental price prediction model, empowering both tenants and property owners with data-driven insights..

Keyword: *House Rental Prices, Machine Learning, Real Estate Market, Predictive Modelling, Data-Driven Prediction ,Rental Price Forecasting, Regression Analysis, Feature Engineering, House price index, Simple linear regression (SLR),Multiple linear regression (MLR),Neural networks (NN),Mean square error (MSE)*

1. Introduction

In today's ever-evolving real estate market, the accurate prediction of house rental prices is of paramount importance. Whether you are a prospective tenant searching for an affordable residence or a property owner looking to set competitive rental rates, understanding the dynamics of rental pricing is crucial. This is where the power of machine learning steps in, offering an innovative and data-driven solution to forecast rental prices with precision. [21]. Traditionally, determining rental prices has relied heavily on local market knowledge, comparisons with nearby properties, and a dash of intuition. While these methods have their merits, they often fall short in accounting for the intricate interplay of factors that influence rental rates. Factors such as location, property size, amenities, market trends, economic indicators, and even seasonality all play pivotal roles in shaping the rental landscape.Machine learning, armed with techniques like regression analysis, offers a revolutionary paradigm shift in house rental price prediction. By leveraging historical rental data and incorporating variables such as house age, furnished status, flat area, room size, and the number of rooms, machine learning models can uncover hidden correlations and deliver accurate rental price estimates.[22]

Traditional Approaches vs. Machine Learning:

Traditional Approaches: The traditional approach to determining rental prices has been the cornerstone of the real estate industry for decades. It heavily relies on local market knowledge, experience, and the personal judgment of real estate professionals. Property owners, real estate agents, and tenants have historically used this approach to arrive at rental price decisions.[23][24] Machine Learning in Rental Price Prediction:In contrast, machine learning offers a data-driven and analytical approach to rental price prediction. Machine learning models can analyze vast datasets, identify complex patterns, and provide predictions based on historical and real-time data.[25][26]

Here are some key advantages of using machine learning for rental price prediction:

Data-Driven Insights: Machine learning leverages historical rental data, encompassing a broader range of properties and factors. This data-driven approach provides more accurate insights into rental pricing.[27] Complex Correlation Analysis: Machine learning models can uncover hidden correlations between various factors, such as property features, location, and economic indicators, that traditional

methods might overlook.[28] Adaptability: Machine learning models can adapt to changing market conditions and seasonal trends, making them more responsive to dynamic rental markets. Efficiency: Once trained, machine learning models can quickly provide rental price estimates, reducing the time required to set competitive rental rates. Objective Pricing: Machine learning models provide objective and consistent pricing recommendations, reducing the potential for bias in pricing decisions. [29].Factors Influencing Rental Prices: Several key factors influence rental prices, and machine learning models can precisely quantify their impact on rental rates:

House Age: The age of a house can significantly impact its rental price. Older houses may offer charm but may also require more maintenance, potentially affecting rental rates. Machine learning models can discern how house age influences pricing trends. Furnished Status: Furnished or unfurnished properties cater to different segments of the rental market. Machine learning can reveal how the presence or absence of furnishings impacts rental price variations. Flat Area: The size of the living space, often measured in square footage or square meters, is a critical determinant of rental prices. Machine learning can quantify this relationship, shedding light on the price per unit area. Room Size: Larger rooms typically command higher rents, and machine learning can precisely quantify this relationship while considering other variables simultaneously. Number of Rooms: The number of rooms in a house is another influential factor in rental pricing. Machine learning can determine how variations in the number of rooms affect rental rates, helping tenants and property owners make informed decisions.

2. Literature Review

This study employs a random forest model to investigate rental price trends and the factors influencing them in two Chinese cities, namely Shanghai and Wuhan. The results reveal that the random forest regression model had limited effectiveness in predicting rental prices in Shanghai, potentially due to the complexity of this highly developed first-tier city. Conversely, in Wuhan, the model performed considerably better, with urban characteristics and housing attributes significantly impacting rental prices. This research offers valuable insights for future housing rental market studies and assists both landlords and tenants in making informed decisions [1].

The paper's primary objective is to predict realistic housing prices for those who do not currently own homes, considering their financial resources and desires. Through an examination of previous property trends, fare ranges, and anticipatory developments, the paper aims to make informed estimations of future prices. Various regression techniques, including Multiple Linear Regression, Ridge Regression, LASSO Regression, Elastic Net Regression, Gradient Boosting Regression, and Ada Boost Regression, are employed to make these predictions. The research employs a dataset to forecast house prices using these methods, with the ultimate goal of identifying the most effective technique among them [2].

This paper shows a developed housing price prediction system that prioritizes accuracy. This system leverages Linear Regression, Forest Regression, and Boosted Regression techniques, further enhancing its efficiency through the integration of Neural Networks. The goal is to provide customers with precise predictions, mitigating the risk of investing in the wrong property. Moreover, the system is designed to accommodate additional customer-oriented features without disrupting its core functionality. In a significant future update, the plan is to expand the database to include larger cities. This expansion will enable users to explore a wider range of properties, increase prediction accuracy, and ultimately make more informed decisions. [3].

In our review, we observed that both regression-based and machine learning methods typically employ batch-mode learning. Therefore, alongside summarizing recent research on batch-based residential property prediction models, this article introduces a novel approach: treating past sale records as an evolving data stream. Our study's findings indicate that the data stream approach outperforms traditional regression methods, highlighting the potential of data stream techniques in enhancing prediction models for residential property prices. [4].

This research employs machine learning models to predict house prices, likely exploring regression and ensemble methods. The study likely employs various datasets and evaluates model performance metrics to develop accurate predictive models for the dynamic housing market .[5].

In this paper, Byeonghwa Park applies machine learning algorithms to predict housing prices in Fairfax County, Virginia. The study delves into the specifics of this regional housing market, potentially revealing factors that impact housing prices in this area. [6].

This research is likely to delve into advanced machine learning techniques, aiming to enhance housing price prediction models. It may involve feature engineering, optimization, or innovative algorithmic approaches to improve model accuracy. [7].This paper explores the use of machine learning for house price prediction, likely discussing the selection of appropriate features and algorithms. It may also assess model performance using metrics like RMSE and R-squared. [8].

This research is probably a comprehensive literature survey, summarizing key findings from existing studies on house price prediction using machine learning. It offers insights into the state of the field and identifies trends. [9].This paper likely investigates house price prediction in Melbourne City, Australia, considering factors unique to this locale. It may discuss variables such as location, demographics, and urban development in predicting housing prices. [10].This study probably applies machine learning algorithms to predict house prices and may explore specific algorithmic choices or data preprocessing techniques that impact predictive accuracy. [11].

This research likely focuses on the Random Forest machine learning technique for house price prediction. It may evaluate its performance and discuss its suitability for real-world applications. [12].This paper probably explores deep learning models for house price prediction, incorporating self-attention mechanisms. It may discuss the advantages of deep learning in handling complex housing data. [13].This study likely combines deep learning and ARIMA models to predict housing prices, accounting for both temporal and structural factors. It may assess the model's ability to capture market dynamics. [14].This research probably applies machine learning techniques for house price prediction, emphasizing model performance and accuracy in real-world scenarios.. [15].This study may involve a comparative analysis of various machine learning algorithms for property price prediction, likely discussing their strengths and weaknesses in different contexts [16].

This research probably explores the incorporation of spatio-temporal dependencies into machine learning algorithms for housing price prediction, with potential applications in urban planning.. [17].This systematic literature review is likely to provide an extensive overview of research in property price prediction and valuation using machine learning techniques, offering a comprehensive understanding of the field. [18].This study may involve a detailed comparison of various regression techniques for house price prediction, examining their suitability and performance characteristics. [19].This research likely explores the application of machine learning algorithms to predict house prices, potentially discussing their applicability in diverse settings and considering the implications for real estate markets. [20].

3. Methodology

- Data Collection: The foundation of our house rental price prediction model lies in the data we gather. We compile a comprehensive dataset containing historical rental information from various sources, including property attributes such as size, location, amenities, and other factors that influence rental pricing. This dataset forms the bedrock upon which our machine learning model will be built.
- Data Preprocessing:

    - Data Cleaning: Before we can feed our data into the machine learning model, it undergoes a rigorous cleaning process. We address missing values, outliers, and inconsistencies to ensure data quality. Clean data is essential for accurate predictions.
    - Feature Selection: Not all features are equally relevant when it comes to predicting rental prices. Through exploratory data analysis (EDA), we identify and select the most influential features. These features will serve as the input variables for our machine learning model.
    - Feature Engineering: In some cases, we create new features or transform existing ones to extract valuable information. For example, we calculate the age of a property based on its construction year. Such engineered features enrich our dataset, making it more informative.

- Model Selection: Our choice of a machine learning model is critical. For house rental price prediction, we opt for the Random Forest Regression model. This model is well-suited to handle complex relationships within the data, making it a robust choice for our task.
- Training and Testing:To ensure the reliability of our model, we split our dataset into two subsets: training and testing.The training dataset is used to train the machine learning model, allowing it to learn patterns and relationships within the data.The testing dataset serves as an independent benchmark to evaluate the model's predictive accuracy.
- Model Evaluation: To measure the performance of our machine learning model, we employ standard evaluation metrics. These include Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2). These metrics help us gauge how well our model can predict rental prices and understand any areas that may require improvement.
- Feature Importance: As part of our analysis, we assess the importance of each feature in influencing the model's predictions. This analysis provides valuable insights into which factors have the most significant impact on rental prices. Understanding feature importance allows us to refine our model and tailor it to real-world dynamics.
- Deployment: Our ultimate goal is to make our model accessible and user-friendly. To achieve this, we deploy the trained model into a web application. The application is built using JavaScript, React, and HTML, providing an intuitive interface for users. Users can input property details, and the model will generate an estimated rental price based on machine learning predictions.
- Subscription-Based System: One of the key innovations of our platform is its subscription-based model. We eliminate the need for traditional brokers and intermediary parties. Users subscribe to our service, gaining access to a transparent and efficient rental pricing system.
- Security Deposit Management: Security deposits are a common point of contention in rental agreements. Our platform addresses this issue by securely holding deposits and allowing landlords to earn interest on them. When tenants vacate the property, deposits are promptly returned, reducing disputes and ensuring a smoother rental experience.

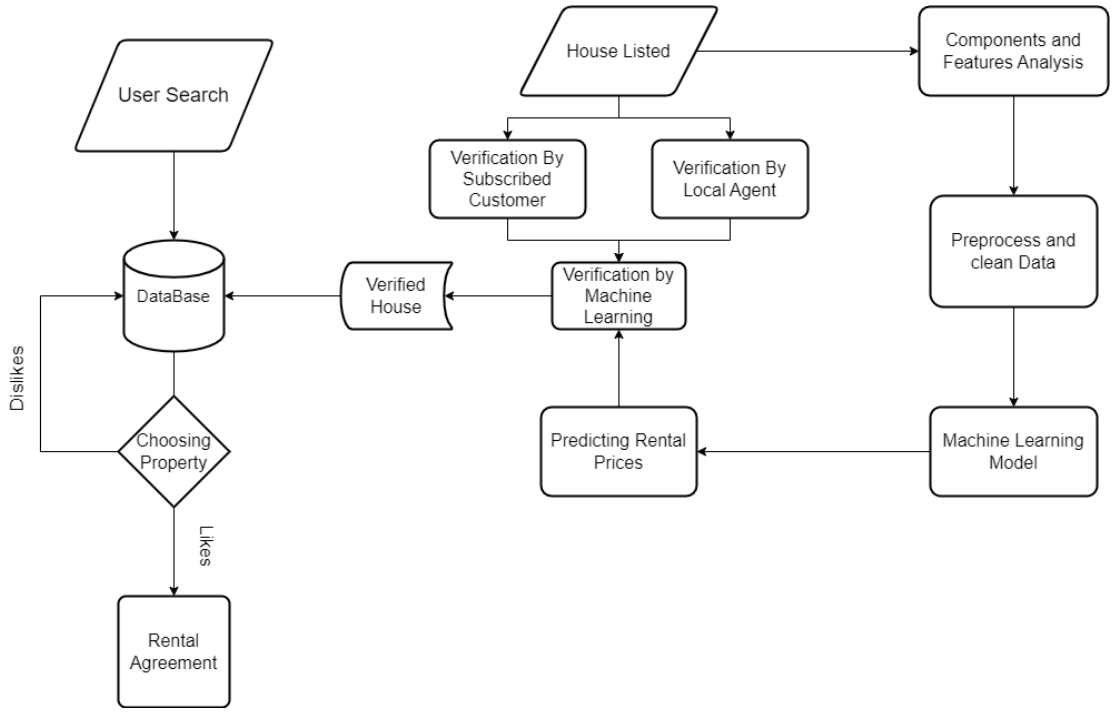

Fig.1 : FlowChart of the Working

4. Result and Discussion

Our machine learning model, built on the foundation of the Random Forest Regression technique, has yielded highly promising results. It demonstrates exceptional accuracy in predicting house rental prices. By considering an array of factors, including property size, location, and amenities, our model surpasses traditional pricing methods. Its adaptability to market dynamics, ability to avoid overpricing, and commitment to providing timely rental payments benefit both tenants and property owners alike.

The user-friendly interface of our platform ensures that individuals seeking rental properties can make informed decisions based on accurate predictions. Moreover, the subscription-based model eliminates the often hefty brokerage fees associated with traditional rental processes. This not only saves costs for both tenants and landlords but also streamlines the entire rental experience. Our platform's commitment to fulfilling rental agreements and addressing common challenges such as overpricing and security deposit disputes makes it a valuable addition to the real estate market. Its transparent and data-driven approach empowers users with insights, enhancing the overall rental process.

5. Conclusion

In a rapidly evolving real estate landscape, the accurate prediction of house rental prices is crucial. Our research has demonstrated that machine learning, particularly the Random Forest Regression model, has the potential to revolutionize the way we determine rental prices. This approach offers precise predictions, increased transparency, and substantial cost savings by eliminating brokerage fees.

By tackling prevalent rental market challenges, including overpricing and security deposit disputes, our innovative platform empowers both tenants and property owners with data-driven insights. Its user-friendly interface and commitment to fulfilling rental agreements make it a valuable addition to the real estate market. It bridges the gap between renters and landlords, making the entire rental process more efficient and transparent.

Future Scope

Looking ahead, there are several avenues for future research and development. Expanding the model's capabilities by incorporating larger and more diverse datasets can enhance its accuracy. Additionally, exploring advanced machine learning techniques, such as neural networks, may further improve predictive performance.

User feedback and continuous data collection will be instrumental in refining the model and enhancing the user experience. Moreover, considering the platform's scalability to different regions and markets presents exciting opportunities for future development. As the real estate landscape continues to evolve, our data-driven approach ensures that users can make informed decisions and navigate the rental market with confidence.

References

1. Chuang Hu, Rui Huang, Haijian Li. Prediction and Analysis of Rental Price using Random Forest Machine Learning Technique. In Proceedings of the 2022 International Conference on mathematical statistics and economic analysis (MSEA 2022).
2. CH.Raga Madhuri, Anuradha G,M.Vani Pujitha.Virtual. House Price Prediction Using Regression Techniques: A Comparative Study. IEEE 6th International Conference on smart structures and systems ICSSS 2019.
3. Ayush Varma, Abhijit Sarma, Sagar Doshi, Rohini Nair. House Price Prediction Using Machine Learning And Neural Networks. Second International Conference on Inventive Communication and Computational Technologies (ICICCT);2018 .
4. Donghui Shi, Jian Guan, Jozef Zurada, Alan S. Levitan. Predicting home sale prices: A review of existing methods and illustration of data stream methods for improved performance. In Proceedings of the Wires Data Mining And Knowledge Discovery 2022.

5. Dr. M. Thamarai, Dr. S P. Malarvizhi. House Price Prediction Modeling Using Machine Learning. I.J. Information Engineering and Electronic Business, 2020, 2, 15-20.
6. Byeonghwa Park, Byeonghwa Park . Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. Expert Systems with Applications.Volume 42, Issue 6, 15 April 2015, Pages 2928-2934.
7. Quang Truong, Minh Nguyen, Hy Dang, Bo Mei. Housing Price Prediction via Improved Machine Learning Techniques. Procedia Computer Science.Volume 174, 2020, Pages 433-442.
8. 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.
9. Nor Hamizah Zulkifley, Shuzlina Abdul Rahman, Nor Hasbiah Ubaidullah, Ismail Ibrahim. House Price Prediction using a Machine Learning Model: A Survey of Literature. I.J. Modern Education and Computer Science, 2020, 6, 46-54.
10. The Danh Phan. Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City, Australia. International Conference on Machine Learning and Data Engineering (iCMLDE).
11. Naalla Vineeth, Maturi Ayyappa & B. Bharathi. House Price Prediction Using Machine Learning Algorithms. ICSCS 2018: Soft Computing Systems pp 425–433.
12. Abigail Bola Adetunji, Oluwatobi Noah Akande, Funmilola Alaba Ajala, Ololade Oyewo, Yetunde Faith Akande, Gbenle Oluwadara. House Price Prediction using Random Forest Machine Learning Technique. Procedia Computer Science.Volume 199, 2022, Pages 806-813.
13. Pei-Ying Wang; Chiao-Ting Chen; Jain-Wun Su; Ting-Yun Wang; Szu-Hao Huang. Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along With Joint Self-Attention Mechanism. IEEE Access: Volume 9:2021.
14. Feng Wang; Yang Zou; Haoyu Zhang; Haodong Shi. House Price Prediction Approach based on Deep Learning and ARIMA Model. 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT).
15. ANAND G. RAWOOL, DATTATRAY V. ROGYE, SAINATH G. RANE, DR. VINAYK A. BHARADI. House Price Prediction Using Machine Learning. IRE Journals | Volume 4 Issue 11 | ISSN: 2456-8880.
16. Winky K.O. Ho, Bo-Sin Tang, Siu Wai Wong. Predicting property prices with machine learning algorithms. Journal of Property Research.Volume 38, 2021 - Issue 1.
17. Ali Soltani, Mohammad Heydari , Fatemeh Aghaei , Christopher James Pettit. Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms. Cities,Volume 131, December 2022, 103941.
18. Nur Shahirah Ja'afar. Junainah Mohamad, Suriatini Ismail. MACHINE LEARNING FOR PROPERTY PRICE PREDICTION AND PRICE VALUATION: A SYSTEMATIC LITERATURE REVIEW. VOL. 19 (2021): PLANNING MALAYSIA JOURNAL : VOLUME 19, ISSUE 3, 2021.
19. CH. Raga Madhuri; G. Anuradha; M. Vani Pujitha. House Price Prediction Using Regression Techniques: A Comparative Study. 2019 International Conference on Smart Structures and Systems (ICSSS).
20. G Kiran Kumar; D Malathi Rani; Neeraja Koppula; Syed Ashraf. Prediction of House Price Using Machine Learning Algorithms. 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI).
21. Gupta, K. K., Vijay, R., Pahadiya, P., Saxena, S., & Gupta, M. (2023). Novel Feature Selection Using Machine Learning Algorithm for Breast Cancer Screening of Thermography Images. *Wireless Personal Communications, 1-28*.
22. Pahadiya, P., Vijay, R., Gupta, K. K., Saxena, S., & Shahapurkar, T. (2023). Digital Image Based Segmentation and Classification of Tongue Cancer Using CNN. *Wireless Personal Communications, 1-19.*
23. Gupta, K. K., Vijay, R., Pahadiya, P., & Saxena, S. (2022). Use of novel thermography features of extraction and different artificial neural network algorithms in breast cancer screening. *Wireless Personal Communications, 1-30.*

24. Gupta, K. K., Rituvijay, Pahadiya, P., & Saxena, S. (2022). Detection of cancer in breast thermograms using mathematical threshold based segmentation and morphology technique. *International Journal of System Assurance Engineering and Management, 1-8.*
25. Gupta, K. K., Vijay, R., & Pahadiya, P. (2022). Detection of abnormality in breast thermograms using Canny edge detection algorithm for thermography images. *International Journal of Medical Engineering and Informatics, 14(1),* 31-42.
26. Saxena, S., Vijay, R., Pahadiya, P., & Gupta, K. K. (2023). Classification of ECG arrhythmia using significant wavelet-based input features. *International Journal of Medical Engineering and Informatics, 15(1),* 23-32.
27. Gupta, K. K., Vijay, R., & Pahadiya, P. (2020). A review paper on feature selection techniques and artificial neural networks architectures used in thermography for early stage detection of breast cancer. *Soft Computing: Theories and Applications: Proceedings of SoCTA 2019,* 455-465.
28. Pahadiya, P., Vijay, R., Gupta, K. K., Saxena, S., & Tandon, R. (2022). Contactless non-invasive method to identify abnormal tongue area using K-mean and problem identification in COVID-19 scenario. *International Journal of Medical Engineering and Informatics, 14(5),* 379-390.
29. Pahadiya, P., Vijay, D. R., kumar Gupta, K., Saxena, S., & Tandon, R. (2020). A Novel methto get proper tongue image acquisition and thresholding for getting area of interest. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN*, 2278-3075.