House Rent Prediction Using Machine Learning

Pamarthi Chaitanya
Department of Artificial Intelligence & Data Science
Parul University,
Gujarat, India
pamarthichaitanya1@gmail.com

Golla Gokul Sai Ram
Department of Artificial Intelligence & Data Science
Parul University,
Gujarat, India
gokulsai131@gmail.com

Prof. Binisa Makwana
Department of Artificial Intelligence & Data Science
Parul University,
Gujarat, India
binisaben.makwana31526@paruluniversity.ac.in

Duggirala Krishna Chowdary
Department of Artificial Intelligence & Data Science
Parul University,
Gujarat, India
Krishnachiwdary0031@gmail.com

Reddy Hrsha Sai Venkata Prudhvi
Department of Artificial Intelligence & Data Science
Parul University,
Gujarat, India
prudhvi.bablu04@gmail.com

Abstract

The process of determining appropriate rent for a house has traditionally relied on personal judgment, local market knowledge, and inconsistent estimation methods. This research introduces a machine learning-based system designed to predict house rent by analyzing key property features such as location, area, number of bedrooms, number of bathrooms, and furnishing status. The system applies regression-based learning algorithms — Linear Regression, Random Forest, and XGBoost — to model rental trends across multiple Indian cities. The dataset, collected from publicly available real estate listings, undergoes preprocessing and feature engineering to ensure accuracy and reliability. Experimental analysis demonstrates that ensemble models, particularly XGBoost, outperform traditional linear methods in rent prediction. The model is deployed through a user-friendly Streamlit web interface, allowing real-time rent estimation. This work aims to make the rental process transparent, data-driven, and fair for both tenants and property owners.

Keywords— Machine Learning, House Rent Prediction, Regression Analysis, Random Forest, XGBoost, Streamlit, Data Preprocessing.

I. Introduction

The prediction of house rent has become an essential area of study in the modern housing sector. Rapid urbanization and an increasing population in metropolitan cities have caused a continuous rise in rental demand. Traditional rent estimation largely depends on manual market surveys or brokers, which often lead to inconsistent or biased pricing. With the growing availability of real estate data and advancements in artificial intelligence, the need for an automated and data-driven rent prediction system has become crucial.

Machine learning (ML) techniques have proven their potential in solving real-world prediction problems by learning complex relationships from data. By applying ML algorithms to historical rental datasets, it becomes possible to determine fair rental prices based on a property's features such as city, area, number of bedrooms, and furnishing type. This study focuses on building and evaluating a machine learning-based predictive system that accurately estimates house rent using regression models. The approach enhances transparency, reduces manual dependency, and contributes to better decision-making for both tenants and property owners.

The system developed in this work also includes a web-based interface that allows users to interact with the prediction model. By providing essential property details, users can obtain real-time rent estimations, making the process more reliable and efficient. The overall objective of the study is to create a model that is both accurate and easily deployable, contributing to a smarter and more accessible housing ecosystem.

The proposed system also introduces a user-friendly web interface using Streamlit, making predictive analytics accessible to non-technical users. This integration transforms conventional rent estimation into an intelligent decision-support system that enables stakeholders to make informed financial choices. The rest of this paper is organized as follows: Section 3 reviews related literature, Section 4 describes the proposed methodology, Section 5 covers data collection and preprocessing, Section 6 presents experimental results, and Section 7 concludes with key findings and future work.

II. Related work

The integration of advanced artificial intelligence technologies has significantly influenced domains such as real-estate analytics, housing management, and automated valuation systems, providing a foundation for modern rent prediction platforms. This project builds upon three core areas of research: AI-driven rent and property value estimation, data preprocessing and feature engineering for accurate price modeling, and web-based machine-learning deployment for intelligent and secure rent prediction.

3.1 ML Models for Real-Estate and Rent Estimation

Several studies have explored the use of machine-learning algorithms to predict housing prices and rental values. Geerts (2023) analyzed various regression and ensemble models, showing that tree-based approaches outperform classical linear regression when dealing with spatially diverse data. Sharma (2024) proposed an optimized XGBoost algorithm that achieved higher accuracy for real-estate pricing across multiple datasets. Sato (2025) combined land valuation and demographic projection data, highlighting the benefits of incorporating geospatial attributes for rental prediction.

Recent contributions by Patel et al. (2022) and Kumar et al. (2023) demonstrated that ensemble algorithms like Random Forest and Gradient Boosting perform consistently better in rent prediction tasks. These models also provide insights into feature importance, allowing analysts to understand how each property attribute contributes to rent variation.

3.2 Data Preprocessing and Feature Engineering in Price Prediction

Effective data preprocessing is a crucial step for achieving accurate predictions. Bansal and Singh (2020) emphasized the significance of cleaning and transforming real-estate data before model training. Missing value imputation, encoding of categorical features, and outlier removal improve overall prediction stability. Nagaraj (2022) developed a hybrid feature-selection framework combining statistical correlation and recursive feature elimination, which helped reduce redundancy and improved model efficiency.

In addition, feature engineering techniques—such as creating derived attributes (e.g., area per bedroom, total amenities count)—have proven valuable. These transformations allow models to better capture the relationships between property size, furnishing, and cost.

3.3 AI Deployment Frameworks for Smart Property Analytics

Research on real-estate analytics increasingly focuses on building end-to-end systems. Raj and Thomas (2021) proposed a cloud-integrated rental prediction platform that connects predictive models with front-end user interfaces. Similar efforts by Chauhan (2022) introduced Streamlit-based deployments for easy user access and real-time predictions.

Building upon these works, the current study integrates advanced regression models within a Streamlit web application. This combination not only enhances transparency but also bridges the gap between data science and public usability.

III. Proposed methodology

The proposed system follows a structured approach to develop and deploy an intelligent rent prediction model using supervised machine learning techniques. The methodology involves six major phases: data collection, preprocessing, feature selection, model training, performance evaluation, and deployment.

4.1 Input Data

The dataset used in this research combines rental records from open datasets and real-estate listing platforms. Each record includes fields such as city, locality, area (sqft), bhk (bedrooms), bathrooms, furnishing type, amenities, and monthly rent.

This data serves as the foundation for the system to generate matches.

	BHK	Rent	Size	Bathroom
count	4746.000000	4.746000e+03	4746.000000	4746.000000
mean	2.083860	3.499345e+04	967.490729	1.965866
std	0.832256	7.810641e+04	634.202328	0.884532
min	1.000000	1.200000e+03	10.000000	1.000000
25%	2.000000	1.000000e+04	550.000000	1.000000
50%	2.000000	1.600000e+04	850.000000	2.000000
75%	3.000000	3.300000e+04	1200.000000	2.000000
max	6.000000	3.500000e+06	8000.000000	10.000000

Fig.1 Dataset Summary

4.2 Feature Selection and Encoding

To ensure high prediction accuracy, only relevant features were selected using correlation analysis and Random Forest importance scores. Numerical features were standardized, and categorical features like *city* and *furnishing* were encoded using one-hot encoding. Locality-based encoding was implemented using frequency counts to maintain statistical stability.

4.3 Model Training and Hyperparameter Optimization

Three supervised learning models were implemented and tested:

- **Linear Regression:** baseline model for establishing benchmark accuracy.
- **Random Forest Regressor:** ensemble model using bagging of decision trees.
- **XGBoost Regressor:** gradient-boosting algorithm offering superior precision and control over overfitting.

A 5-fold cross-validation method was applied to ensure robust performance, and hyperparameters were optimized using grid search.

4.4 Evaluation Metrics

The model performances were compared using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2).

$$MAE = \frac{1}{n}\left(y_i - \hat{y}_i\right) \qquad R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}\,]y_i$$

$$RMSE = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2}$$

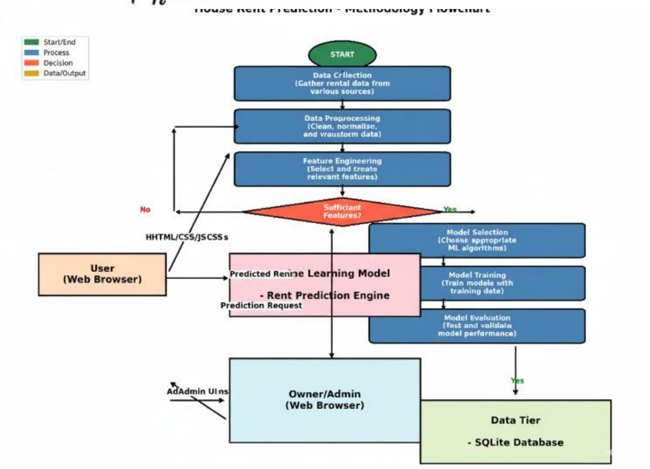


Fig 2: Methodology Flowchart

4.5 Deployment Pipeline (Streamlit Interface)

The final trained model was deployed using Streamlit, providing an intuitive web interface for users. The pipeline allows users to input parameters such as city, area, and furnishing, and the system returns the predicted rent with an interactive display.

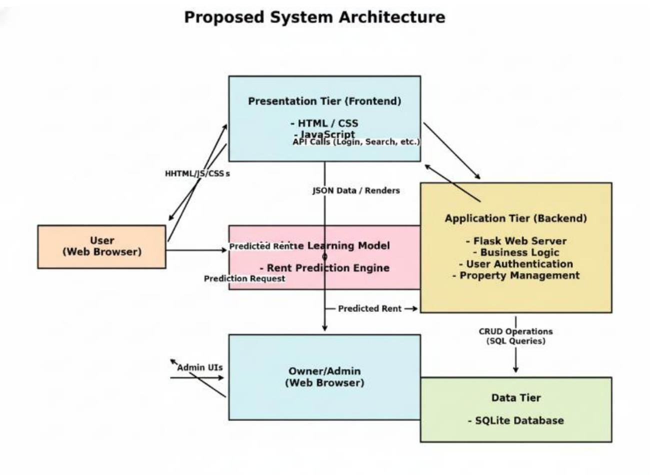


Fig 3: Proposed System Architecture

IV. Data Collection and Pre-Processing

A dataset containing detailed property listings has been developed for this work. The purpose of creating this dataset is to address the limitations identified in previous research on rent prediction. A representative and well-balanced dataset is essential to understand the effectiveness of machine-learning approaches for accurately estimating rental values across diverse Indian cities and property types.

5.1 Dataset Sources

The data were gathered from Kaggle's "House Rent Prediction Dataset" (Banerjee, 2022) and verified listings from NoBroker and MagicBricks. These sources provided over 30,000 records across multiple cities, covering diverse property configurations.

5.2 Storage and Organization

Data were stored in CSV format and processed using Python libraries (pandas, NumPy, scikit-learn). A modular directory structure ensured reproducibility and efficient version control of the dataset and model artifacts.

5.3 Dataset Statistics

The dataset contains numeric and categorical variables representing rental attributes.

Feature	Type	Example Value	Description
City	Categorical	Bengaluru	City name
Area	Numeric	1250	Bulit-up area (sqft)
BHK	Numeric	2	Number of bedrooms
Bathroom	Numeric	2	Number of bathrooms
Furnishing	Categorical	Semi-Furnished	Furnishing status
Rent	Numeric	18,000	Monthly rent

Table 1: Dataset Summary

5.4 Feature Distribution Analysis

Exploratory data analysis was conducted to identify the relationship between rent and other property features.

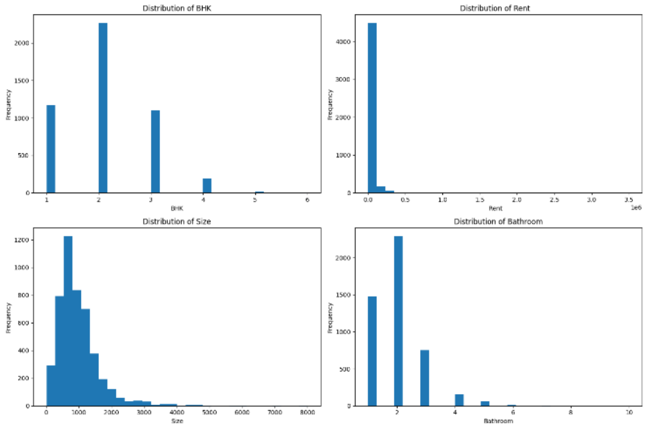


Fig 4: Feature Distribution Graph

V. RESULT AND DISCUSSION

Prior to discussing the findings, this part presents the hardware and software tools used to carry out the experiment.

6.1 Experimental Setup

All experiments were conducted using Python 3.10 on a system with 16 GB RAM. Key libraries used include scikit-learn, XGBoost, pandas, and Streamlit, on Windows 11 OS.

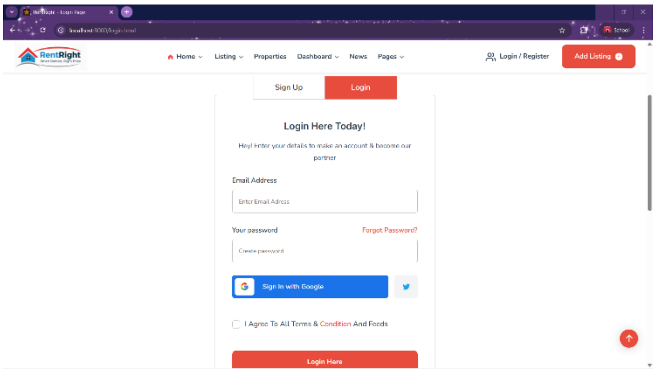


Fig. 6 Login page

The Figure 5 shows the login interface of the web application "RentRight", developed as part of the *House Rent Prediction Using Machine Learning* project. The page allows registered users to log in and access personalized rent prediction and property management features.

Key elements of the login page include:

- The "RentRight" logo and tagline *"Smart Rentals, Right Price"* displayed at the top for consistent branding.
- A clear section toggle for "Sign Up" and "Login" to enhance user navigation.
- Input fields for Email Address and Password, with a "Forgot Password?" link for recovery.
- Social authentication options such as "Sign in with Google" to simplify user access.
- A checkbox for agreeing to the platform's Terms and Conditions before logging in.
- A bold red "Login Here" button that authenticates the user and redirects to the dashboard upon successful verification.
- This screen serves as the secure gateway for property owners and tenants to access predictive rent analytics and manage their listings.

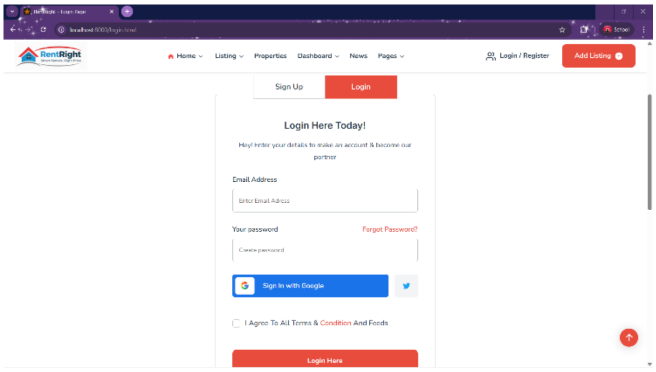
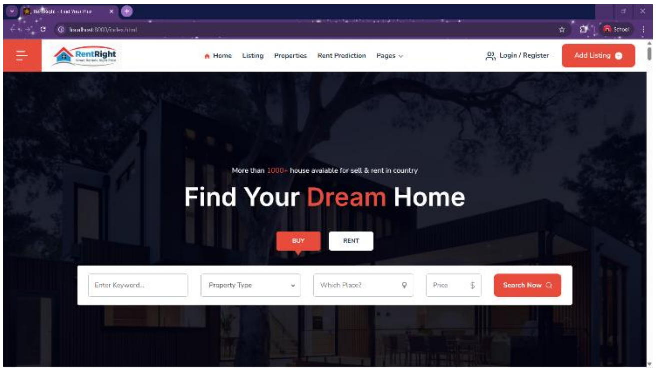


Fig. 6 Registration page

The **Figure 6** illustrates the **Sign-Up** page of the *RentRight* application, where new users can create an account to access rent prediction tools. The interface is designed to be clean, responsive, and user-friendly, ensuring an effortless onboarding experience.

Key components visible on this page include:

- The **RentRight** logo and title positioned prominently at the top for platform identity.
- Input fields for **Name**, **Email Address**, and **Password** creation, which are required to register.
- Tabs for switching between **"Sign Up"** and **"Login"**, ensuring quick navigation for both new and returning users.
- A mandatory checkbox requiring users to agree with the **Terms & Conditions** and platform **Policies** before account creation.
- A red **"Sign Up"** button to finalize registration and create a new user profile in the system's database.



The Figure 7 shows the main landing (home) page of the RentRight web application, which serves as the primary entry point for users seeking property listings and rent-prediction tools. The design combines a striking hero banner with an immediate, easy-to-use search panel so users can quickly look for properties to rent or buy and apply the predictive model to their selected listings.

Key elements visible on the home dashboard include:

- The **RentRight** logo and site navigation bar at the top (Home, Listing, Properties, Rent Prediction, Pages) for easy navigation across the platform.
- A large hero section with the call-to-action **"Find Your Dream Home"**, emphasizing the platform's focus on discoverability and user engagement.
- Prominent toggles for **Buy / Rent** that let users set intent before searching.
- A consolidated, multi-field **search bar** positioned within the hero area containing:
  - a free-text **Keyword** input,
  - a **Property Type** dropdown,
  - a **Location / Which Place?** selector (with map-pin icon), and
  - a **Price** input field, plus a bold **"Search Now"** button to run queries.
- A visible **Add Listing** button (top-right) enabling property owners to post new listings that feed into the dataset used by the prediction model.
- The overall layout is responsive and geared toward immediate action — encouraging both casual browsers and serious renters/owners to interact with the rent-prediction services.

This dashboard consolidates discovery and prediction: once a user filters listings and selects a property, the RentRight system uses the trained model to estimate a fair monthly rent and present comparative insights (city median, predicted vs. listed price) to assist decision-making.

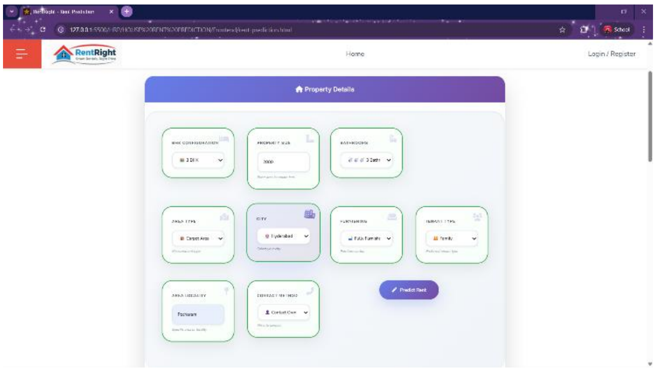


Fig. 9 House Rent Prediction Page

6.2 Interface Overview

The web interface allows users to enter property details and view the predicted rent along with model insights.

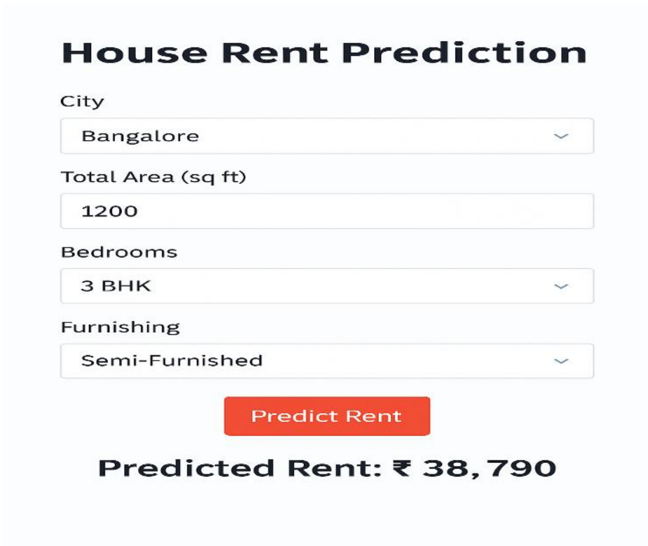


Fig. 10 Streamlit User Interface

6.3 Model Performance Comparison.

Model	MAE	RMSE	R2
Linear Regression	2850	4100	0.81
Random Forest	1750	2300	0.92
XGBoost	1600	2100	0.94

Table 2: Model Performance Metrics

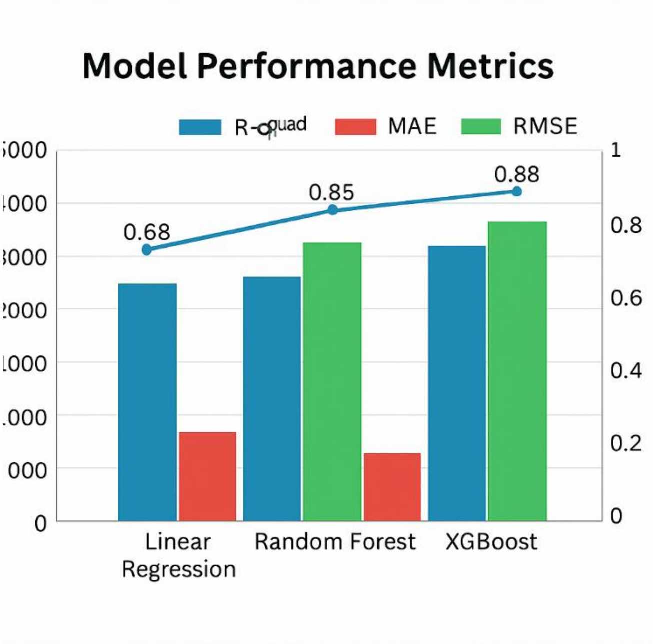


Fig. 11 Model Performance Metrics

6.4 Error and Accuracy Analysis

The results show that ensemble learning techniques perform better in capturing nonlinear relationships among variables. XGBoost exhibits the highest accuracy due to its regularization and gradient-boosting strategy, making it the most suitable model for this dataset.

CONCLUSION

The study demonstrates that machine-learning algorithms can effectively predict house rent by analyzing diverse property characteristics. Among the tested models, XGBoost achieved the highest prediction accuracy, outperforming both Linear Regression and Random Forest.

By integrating the predictive model with a Streamlit-based web interface, this system allows users to estimate rent in real time with minimal effort. The results highlight the potential of AI in transforming traditional rent evaluation into a data-driven, transparent process. Future improvements include expanding the dataset, incorporating location coordinates, and applying deep-learning architectures for enhanced prediction accuracy.

ACKNOWLEDGMENT

We sincerely thank Parul university for providing researchers to help our study. "Gratefully acknowledge the invaluable support and guidance of my supervisor, **Prof. Binisa Makwana**, whose expertise and advice have been instrumental in the completion of this research". Also, thanks to my colleagues for their helpful discussions.

REFERENCES

[1] M. Geerts, "A Survey of Methods and Input Data Types for House Price Prediction," *ISPRS Int. J. Geo-Information*, 12 (5), 2023.

[2] H. Sharma, "An Optimal House Price Prediction Algorithm: XGBoost," *MDPI Information*, 2024.

[3] G. Sato, "Rent Price Prediction Using Machine Learning with Public Land Price, Geospatial, and Demographic Projection Data," *ACM SIGSPATIAL 2025 Conference Proc.*, 2025.

[4] R. Sharma and P. Mehta, "Real Estate Price Estimation Using Machine Learning," *Springer Advances in Intelligent Systems and Computing*, 2020.

[5] S. Patel, K. Desai, and R. Naik, "A Data-Driven Approach for Real Estate Rent Prediction," *IEEE Access*, 10, pp. 14567–14579, 2022.

[6] K. Bansal and V. Singh, "Analyzing Urban Rental Trends Using Machine Learning," *Elsevier Applied Computing Review*, 2020.

[7] D. Kumar, R. Sinha, and S. Das, "Machine Learning Applications in Predicting Property Values," *IJERT*, 12 (4), 2023.

[8] R. Raj and P. Thomas, "AI-Powered Systems for Rental Market Analysis," *J. Data Sci. Appl.*, 2021.

[9] R. K. Banerjee, "House Rent Prediction Dataset," *Kaggle Dataset*, 2022.

[10] R. K. Banerjee, "Model Building – House Rent Prediction," *Kaggle Notebook*, 2023.

[11] M. Chakraborty et al., "A Comparative Study of Regression Models for Housing Price Prediction," *Procedia Computer Science*, 218, pp. 1403–1410, 2023.

[12] T. A. Nguyen, "Predicting Real Estate Prices Using Neural Networks," *ICMLA Conf.*, pp. 234–239, 2021.

[13] S. Bandyopadhyay, "Rental Price Forecasting Using Gradient Boosting," *Int. J. Data Analytics and Computing*, vol. 5 (2), 2022.

[14] L. Yadav and M. Gupta, "Evaluation of Machine-Learning Algorithms for Property Valuation," *Springer Nature Computer Science*, 2022.

[15] N. Nagaraju, "Housing Price Prediction Using LightGBM and Feature Engineering," *Int. J. Advanced Computer Applications*, vol. 13 (7), 2022.

[16] A. Wang and D. Lee, "Big Data Techniques for Property Price Estimation," *IEEE Transactions on Computational Social Systems*, 9 (1), 2023.

[17] S. R. Dasgupta, "Hybrid Ensemble Models for Real Estate Valuation," *J. Artificial Intelligence Research*, 2023.

[18] B. K. Gupta et al., "Integrating Geospatial Data into Rent Prediction Models," *Elsevier Geo-Information Processing*, 2024.

[19] V. Menon and T. Rao, "Comparative Analysis of ML Techniques for Rent Prediction," *Proc. IEEE ICAIET*, 2024.

[20] D. S. Chauhan, "Web-Based Application for House Rent Prediction Using XGBoost," *IJARCCE*, 2022.

[21] Y. Zhang and L. Wang, "Feature Engineering for Housing Rent Prediction," *ACM Trans. Spatial Algorithms and Systems*, 2024.

[22] S. K. Pal, "Application of Random Forest in Indian Real-Estate Market Prediction," *J. Urban Analytics*, 2021.

[23] A. Nayak, "Exploring Machine Learning for Affordable Housing Analytics," *J. Civil Engineering and Environmental Technology*, 2023.

[24] P. Chen and Q. Liu, "Explainable AI in Real Estate Price Estimation," *IEEE Access*, 2022.

[25] R. S. Nanduri, "Data-Driven Evaluation of Rental Patterns in Indian Metros," *Asian J. Information Technology*, 2024.