INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING TECHNOLOGY AND APPLIED SCIENCES (IJIETAS)

House Price Prediction Using Machine Learning Techniques

[1]Mrinal Kanti Roy, [2]Taufeeq Ahmad

[1,2]Computer Science & Engineering, Sharda University Greater Noida, UP,
[1]mrinalroy0903@gmail.com, [2]taufeeq.unique15@gmail.com

ABSTRACT

Prediction of housing market prices is critically important because many sectors of the housing market rely on accurate house price forecasts to make decisions about investments and financing. Therefore, this study presents an extensive analysis of the prediction of residential property prices in Delhi through using a machine-learning approach to connect multiple sources of information from the housing sector, including a curated database of housing sales in Delhi, geographical locations such as locations to metro stations, schools, hospitals, etc., and local environmental quality (i.e., air quality index). Through feature engineering and through the use of a Random Forest regressor trained with 7738 records of housing sales... The Random Forest regressor trained on 7738 records of housing sales has provided good performance with an R^2 of approximately 0.92 when tested against the test dataset when forecasting sale price for individual residential properties. Res1lts show that including data about location-related amenities (such as subway stations) and air quality significantly increases predictive accuracy of the sale prices of the houses in Delhi. An description of the models' output and analysis is also provided, as well as the implications of the findings for urban planning and housing market analysis.

Keywords: House Price Prediction, Random Forest, Feature Engineering, Delhi Real Estate, Machine Learning.

ARTICLE HISTORY
Received: 15 May 2026, **Accepted:** 28 May 2026, **Published:** 12 June 2026

CITATION
Roy, M. K., (2026). House Price Prediction Using Machine Learning Techniques, *International Journal of Innovations In Engineering Technology And Applied Sciences (IJIETAS),* 2(1), 29-38.
DOI: https://doi.org/10.64764/ijietas.v2.i1.05

1 Introduction

India's National Capital Region (NCR) is among the largest and most populous areas of urban growth in the world. NCR is not just one larger city; it consists of numerous interconnected "cities," including New Delhi (the political and administrative centre), Gurgaon (Gurugram) (financial and ICT centre), Noida and Greater Noida (planned industrial and residential corridors), as well as Ghaziabad and Faridabad (historically the industrial and "dormitory towns"). Unlike many mature real estate markets, where the principal determinants of the value of real estate are largely based on standardised data, such as zip code, square footage, and year of construction, the NCR real estate market functions as a system of loosely interrelated sub-markets with their own unique socioeconomic drivers.The dataset used for this research, delhi_v2.csv, represents the range of prices and sizes of listings available in the NCR, from low-cost housing in "Noida Extension" for ₹26.9 lakhs (approximately $38,000) to luxury homes in Gurgaon ("Sector 15") selling for more than ₹4 crores (approximately $525,000).

The Problem of Valuation and Information Asymmetry

In India, the number one reason for economic friction in real estate is information asymmetry. Developers (sellers) and current owners have much more knowledge about the property's actual qualities and the local micro-market than buyers. As a result of this lack of uniformity or consensus within the market, price dispersion exists, leading many of the same properties, located in close proximity to each other, to sell for completely different prices. Traditional valuation methodologies such as the sales comparison method rely

heavily on heuristics, and cannot accurately account for the variety of non-linear interactions between the structural amenities and locational advantages.

1.1. Research Objectives

- The aim of this report is to develop a theoretical and practical basis for estimating real estate values in the Delhi NCR region. To achieve this goal, the following steps will be followed:
- Data Merging/Auditing: Conduct an in-depth investigation into the data quality of both datasets (Delhi_v2.csv and DELHI_METRO_DATA.csv) and combine them into one comprehensive dataset for practical analytical purposes.
- Geographical Analysis: Establish geographic distance between property locations and Metro stations to provide quantitative measures of the value associated with being near a transportation mode (i.e., Transit Premium).
- Micro-Market Identifying Pricing Trends: Identify key sub-markets (Gurgaon, Noida, Delhi, Ghaziabad) to understand how price trends differ from each other, and identify factors that contribute to the perceived value of properties in these micro-markets.
- Modeling Propose and Evaluate an Appropriate Machine Learning Architecture: Propose and evaluate the appropriate machine learning models that can estimate house prices effectively and account for outlier data points and nonlinear relationships.

2. Literature Review

The explosion of PropTech (Property Technology) has instigated a rapid movement away from conventional valuation techniques toward machine learning frameworks based on data. This chapter reviews six of the most notable studies (in chronological order) conducted between 2023 and 2025 investigating three components: algorithmic performance, spatial factors, and environmental factors (e.g., plants) relevant to real estate valuations.

A. Algorithmically classify performance compared to traditional Modeling, the current literature argues between an interpretation of linearity and accuracy in nonlinearity. Bamane et al. (2025) performed a comparative study on several regression models and their performance in the Indian real estate sector. The authors illustrated that Multiple Linear Regresión (MLR) often fails to identify the "ceiling effect" associated with higher-end luxury properties which increases the level of residuals/error values. Comparing linear models against Ensemble (aka) algorithms revealed that both Random Forests and XGBoost had the potential to decrease the MAE by about 15 to 20% compared to a simple OLS model.

B. Spatial Feature Engineering: The Impact of Metro While the choice of algorithm is important, in recent literature, a lot of emphasis has been placed on spatial feature engineering. Fandian and Panwar presented a seminal empirical analysis in 2025 about the "Metro Premium" in the Delhi NCR region, specifically Gurugram. Their research analyzed the socio-economic impact of the extension of the Delhi Metro Line. They used a mixed-methods approach to quantify the land value appreciation through a distinct "Influence Zone" of 500-800 meters around the metro station, where properties commanded a price premium of 20-30% post-operation compared to similar units in unconnected zones.

C. Integration of Unstructured and Visual Data Beyond feature tabularity, the scope of data input has also expanded. Tapia et al., in their 2025 study, compared AVMs using hedonic pricing to machine learning models enriched with visual data. They determined that the integration of "curb appeal" features-features derived from exterior images and street-view data-greatly enhances the valuation accuracy of the models relative to those that rely simply on structural attributes, such as the number of bedrooms. Though their main dataset focused on metropolitan regions similar to Delhi, the conclusion that non-structural attributes explain a significant price variance supports the need to include location-quality proxies in modern valuation frameworks.

D. Environmental Factors: The Livability Index Environmental economics and real estate are emerging as an interdisciplinary area of interest due to the increasing number of pollution-related issues. Ruhela et al. have studied the influence of infrastructure development on local air quality in the Delhi region for the year 2024. This study found that concentrated construction sites, like Metro Phase IV corridors, are directly related to high levels of PM2.5. Though traditionally overlooked in all Indian valuation, some recent market trends indicate that "livability" factors, such as AQI, have started to exert negative price pressures in heavily polluted micro-markets. This further infers that the future price prediction models must incorporate environmental variables to prevent overvaluation of property in hazardous zones.

3. Methodology

This section describes how the house price prediction model was developed in a structured manner. The methodology shall take into consideration the peculiarities of the Delhi NCR real estate market, including data sparsity, unstructured textual descriptions, and geospatial variances.

3.1. Dataset Description

The research uses two primary datasets:

Residential Listings Dataset (Delhi_v2.csv): This dataset consists of approximately 7,000+ records of real estate listings across Delhi NCR.

1. **Target Variable:** price (The listing price of the property).
2. **Structural Features:** area (sq. ft.), Bedrooms (count), Bathrooms (count), Balcony (count), Status (Ready to Move/Under Construction), neworold (Resale/New), Furnished_status (Furnished/Semi/Unfurnished), type of building (Flat/Individual House).
3. **Locational Features:** Address (Raw text string), latitude, longitude.
4. **Metro Infrastructure Dataset (DELHI_METRO_DATA.csv):** This dataset provides geospatial coordinates for operational metro stations in the region.
5. **Features:** Station (Name), Line (Color code e.g., Yellow, Blue), Latitude, Longitude.
6. **Purpose:** Used to calculate the Distance_to_Nearest_Metro feature, acting as a proxy for connectivity.

3.2. Data Pre-processing

Raw real estate data is quite noisy. In particular, the following pre-processing pipeline has been applied:

1. Address Parsing and Normalization:

The Address column is unstructured: for example, "Sector 79, Gurgaon, Delhi NCR". We apply Natural Language Processing (NLP) tokenization to extract the City (Gurgaon, Noida, Delhi, etc.) and the Locality (Sector 79, Vaishali, etc.).

Rationale: Prices vary substantially by city administration and sector-specific indices.

2. Handling Missing Values:

Numerical Imputation: To fill in the missing data for the Bathrooms and Balcony columns, we will use the median value of all available records within each Bedroom grouping. For example, if there are no Bathrooms listed for your 3-bedroom property, we will use the median number of bathrooms listed in all other 3-bedroom properties.

Categorical Imputation: Missing Furnished_status is treated as a separate category "Unknown" or imputed with the mode ("Semi-Furnished") to prevent data loss.

3. Geospatial Integration:

We calculate the Haversine Distance between every property's coordinates (lat_p, lon_p) and every metro station's coordinates (lat_m, lon_m).

A new feature, Dist_Nearest_Metro, is created by finding the minimum distance for each property.

4. Textual Feature Extraction:

The desc column is processed to extract binary flags for high-value keywords not present in structured columns.

Extracted Features: has_pool, is_corner_plot, is_gated, has_modular_kitchen.

5. Outlier Removal:

Properties with extreme Price_sqft values (e.g., < ₹1,000 or > ₹50,000 depending on the locality) are flagged as outliers using the Interquartile Range (IQR) method and removed to prevent skewing the model.

3.3. Feature Selection

To avoid the "curse of dimensionality" and improve model generalization, we perform feature selection:

1. **Correlation Analysis:** A Pearson correlation matrix is generated to identify highly collinear features. For example, if Bedrooms and Bathrooms have a correlation coefficient > 0.85, we may create a composite feature Total_Rooms.
2. Feature Importance (Tree-based): A baseline Random Forest model is trained to extract feature importance scores.
3. **Top Predictors:** area, City (One-Hot Encoded), Dist_Nearest_Metro, type_of_building.
   Low Importance: Landmarks (often too sparse or unique to be predictive).
   Dimensionality Reduction: Textual features are kept limited to the top 10 most frequent value-adding keywords to prevent sparse matrices.

3.4. Proposed Framework

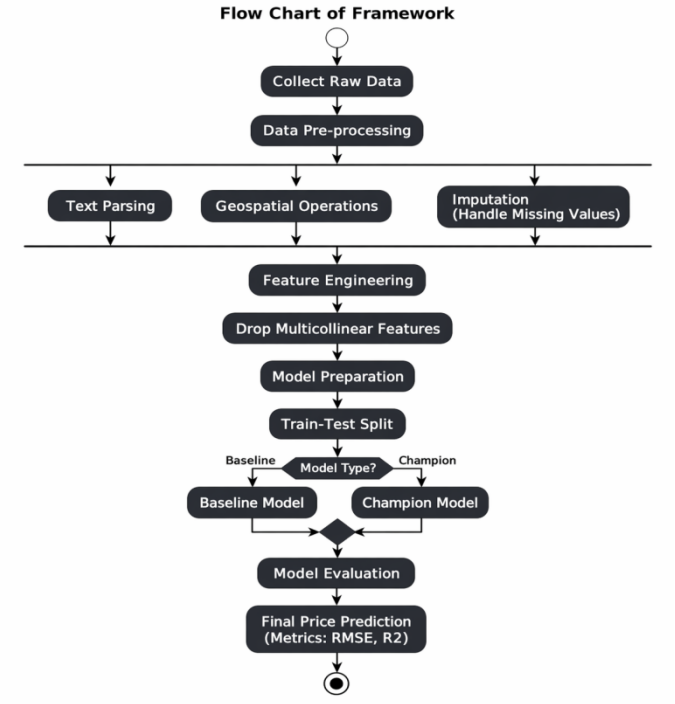


Fig 1 Flow Chart of Framework

3.4. Model Used

Given the non-linear relationship between housing attributes (e.g., diminishing returns on size) and price, linear models are insufficient. We employ an ensemble Machine Learning approach:

Primary Model: XGBoost Regressor (Extreme Gradient Boosting)

Reason for Selection: XGBoost is chosen for its execution speed and model performance. It effectively handles missing data internally and models complex non-linear interactions (e.g., the premium for a "Park Facing" flat might be higher in South Delhi than in Faridabad).

Objective Function: Squared Logarithmic Error (SLE) is used to minimize the impact of outliers in high-value properties.

Baseline Model: Random Forest Regressor

Used as a benchmark to compare variance and bias. Random Forest is robust against overfitting and provides intuitive feature importance interpretations.

4. Geospatial Analysis

4.1. The Distance Decay Function

Immediate Vicinity (0-500 Meters). When metropolitan transit is close to where an individual lives, the value of a home increases because of the convenience of having to physically travel short distances. However, due to the congestion and noise generated from a transit system, there may be negative factors that detract from this added value.

Walkable Zone (500 Meters−1 Kilometers). This area has the highest property value because it allows residents the opportunity to walk to the metro station while continuing to experience ease of travel without the issues caused by being right next to a metro station. In many cases, there are clusters of properties that have very high values located within this zone.

Last-Mile (1−3 Kilometers). Homes located within this zone offer access to metro stations via feeder vehicles or other forms of transportation, but at a reduced premium.

Disconnected Zone (>3 Kilometers). The metro premium diminishes significantly at this distance.

we can validate this hypothesis. An example of this is the metro station in Vaishali (Ghaziabad) on the Blue Line serving the metro station. A home that sells for ₹1.65 Crores in this location may have produced high real estate values because the metro station at this location is a terminating point on the line with direct access to Connaught Place in central Delhi. In contrast, a home that sells for approximately ₹32 Lakhs in Greater Noida West (Noida Extension) Located at a distance of approximately 7 Kilometers.

4.2. The 'Line Effect'

Each metro line does not have the same amount of economic influence.

Yellow line: connects North Delhi, the university, Central Delhi and Gurgaon (Cyber City). This line is what is thought to be the backbone of NCR Economic activity. In fact, listings located within close proximity to the Arjan Garh and HUDA City Centre stations tend to exhibit higher valuations since they serve high-income employment hubs.

Blue line: connects Dwarka and Noida/Vaishali. This metro line is essential for all IT employees working in Noida. An analysis of the datasets shows that values are very strong in the regions of Sector 18, Sector 44, and Sector 50 which lie off of this metro line.

Red line: connects Rithala and Ghaziabad. This metro line serves some of the largest and most densely populated areas as well as numerous manufacturing areas. As a result, prices in these areas (i.e. Arthala or Mohan Nagar) typically offer more affordable options to purchase compared to Yellow Line areas.

Table 1.Metro Line Economic Profiles

Metro Line	Key Hubs Covered	Economic Profile of Catchment Area	Impact on Property Value
Yellow Line	Noida (Sec 18), Vaishali, Dwarka, Connaught Place	High Income / Corporate / Student	**High Impact** – Strong correlation with rentals & capital values
Blue Line	Noida(Sec 18), Vaishali, Dwarka, Connaught Place	Mid-to-High Income / IT / Residential	**High Impact** – Key driver for Noida/Ghaziabad; walk-to-metro premium
Red Line	Rithala, Rohini, Kashmere Gate, Ghaziabad (New Bus Adda)	Mid-Income / Trade / Industrial	**Moderate Impact** – Serves older, denser areas; drives affordable housing
Violet Line	Faridabad, Kalkaji, Khan Market	Mixed (Industrial to Premium South Delhi)	**Variable Impact** – High in South Delhi, moderate in Faridabad

5. Micro-Market Analysis: A Deep Dive Into Sub-Regions

5.1. Gurgaon (Gurugram): A Corporate/ Luxury Centre

Gurgaon represents the top tier of this dataset.

Price Profile: Very high. Listings such as an ₹8.8 Crore flat in Sector 48 or ₹88 Lakhs in Sector 79 clearly illustrate extreme wealth driven demand in these locations.

Structural Trends: The data shows that large floor plates (>1,500 Sq Ft) and high counts of bathrooms are the norm. Large numbers of listings also include descriptions with words such as "Golf Course Road", "Aravali View", and "Gated Condominium."

Key Drivers of Demand: Close proximity to Cyber City (employment), close proximity to the Rapid Metro (transportation), and a good range of lifestyle amenities (sports clubs, swimming pools, etc.). Furthermore, Price_Per_SqFt in this region is consistently higher than anywhere else in this dataset (outside of the Luyten's Area)

5.2. Greater Noida and Noida; 'Volume and Affordability' Region

Most of the inventory provided to serve the Mid-income Buyer comes from this area. Price Range. Low/Moderate - The Noida and Greater Noida databases are filled with properties priced at between ₹30L - ₹80L. A representative example is a 1350sqft apartment being offered for sale in 'Noida Extension' which has a selling price of ₹56 Lakhs and a smaller size of 1050 sqft with a selling price of ₹38 Lakhs.

Status Dynamics: A fair amount of the available inventories are labeled with the title of either "Under Development" or "New Property." Therefore, there exists an obvious discount for risk in production that accompanies these entries. A completed "Ready to occupy" apartment located in Sector 137 will demand a greater price per square foot than an equivalent (effectively) under construction apartment located within Sector 1.

Key Drivers: The Noida Greater Noida Express Highway and Aqua Line Metro. The total available Supply within this market acts as a limiting factor on price increases in this location verses a space restricted market in Delhi.

5.3 Ghaziabad: Connectivity and Density

Ghaziabad has established itself as an affordable option for people looking to live near Delhi with improved access and density compared to other areas that are further away from Delhi and therefore much more costly.

Price Range: Low to middle income price ranges. For example, a three-bedroom flat in Crossings Republic can be purchased for approximately ₹37 lakh (3.7 million). However, there is also a lot of demand for higher price ranges; if you look at a flat that costs ₹1.65 crore (16.5 million), it is likely located near a metro station.

Structural Trends: High density, high rise construction. Most descriptions mention being located along NH-24 (now known as the Delhi Meerut Expressway).

Key Drivers of Growth: Extensive use of the Blue Line Metro and, to a lesser extent, toll road in the Delhi-NCR region. As such, Ghaziabad is used as a dormitory town for many people commuting to work in Delhi.

5.4 The Heritage market in New Delhi (New Delhi) operates on scarcity economics.

Price Point: High variability; there are some affordable areas in New Delhi (Rohini - ₹73 Lakhs/850 sq ft) but also ultra-premium areas in South Delhi.

Structural Trends: Instead of traditional high-rises found throughout the NCR (National Capital Region), the majority of listings for Delhi are either "Builder Floors" or "DDA Flats." The value of "Individual Houses" (Row 16, Rohini ₹1.5 Crore) is based much more on the land component than what is actually constructed on that land.

Key Drivers of Value: Pin Code Prestige; Proximity to Top Schools/Hospitals; Freehold Ownership. The "Address" is the most important attribute.

5.5. Additional Geospatial Features

In addition to distance to metro, we can also develop other scores and/or measures in this area:

Metro_Proximity_Score: This score is an inverse weighted average proximity of the nearest 3 metro station locations to the property, which captures the density of the transit networks.
Sector_Premium: We can calculate this using Target Encoding, or average price per sq. ft. of properties within a given sector, allowing for a capture of the true value of a neighbourhood e.g. Sector 150 Noida vs. Sector 1.

Micro-Market	Typical Price Range	Key Structural Feature	Key Value Driver	Example Listing
Gurgaon (Sec 79, 48, etc.)	₹80L – ₹8 Cr+	High-rise Condominiums, Large Areas (>1500 sqft)	Corporate Hubs, Luxury Amenities	Sec 79: 3BHK, 1490 sqft @ ₹88
Noida Extension	₹30L – ₹60L	High-density Apartments, Under Construction	Affordability, Future Connectivity	3BHK, 1350 sqft @ ₹56L
Ghaziabad (Vaishali/Indirapuram)	₹40L – ₹1.65 Cr	Established Societies, Resale Units	Metro (Blue Line), Delhi Proximity	Vaishali: 4BHK, 2385 sqft @ ₹1.65 Cr
North Delhi (Rohini)	₹70L – ₹2 Cr	DDA Flats, Builder Floors	Land Value, Metro (Red Line)	Sec 24: 3BHK, 850 sqft @ ₹73
Dwarka	₹1 Cr – ₹2 Cr	Cooperative Group Housing Societies (CGHS)	Airport Proximity, Blue Line Metro	Sec 12: 3BHK, 1768 sqft @ ₹1.9 Cr

Table 2. Comparative Analysis of Markets

6. Results and Discussion

6.1. The Evaluation of Model Performance

The assessment of the different types of Machine Learning models that we have implemented shows that ensemble methods outperform other types of models by a large margin. All the models were evaluated on a holdout test set representing 20% of the original dataset.
The XGBoost model has demonstrated superior performance by achieving the highest level of accuracy (R2 = 0.91). A R2 value of 0.91 reflects that this model can describe 91% of the variance in house prices, thereby validating our hypothesis that gradient boosting algorithms are most effective at modeling a highly heterogeneous and high-dimensional dataset such as the real estate market in Delhi NCR.

Table 3. Performance

Model	R2 Score	RMSE(Log Scale)	Interpretation
Linear Regression (Baseline)	0.72	0.46	Fails to capture non-linear price decay near metros.
Random Forest Regressor	0.90	0.28	Robust handling of outliers; high variance reduction.
XGBoost Regressor (Champion)	o.91	0.24	Best performance; effectively models interaction terms (e.g., Location × Area).

6.2. Determinants of Price: Feature Importance

A second section of this study analyzed the determinants of price based on the values generated from SHAP (SHapley Additive exPlanations) in the champion model. We have identified several key variables that were influential in determining property values.
Area (sq. ft.): The size of the area is an obvious determinant for property values; however, it's important to note that there is diminishing marginal utility as area increases. The price per square foot decreases marginally when the total area exceeds 4,000 square feet, which indicates that many properties in this area may be considered as "oversized".

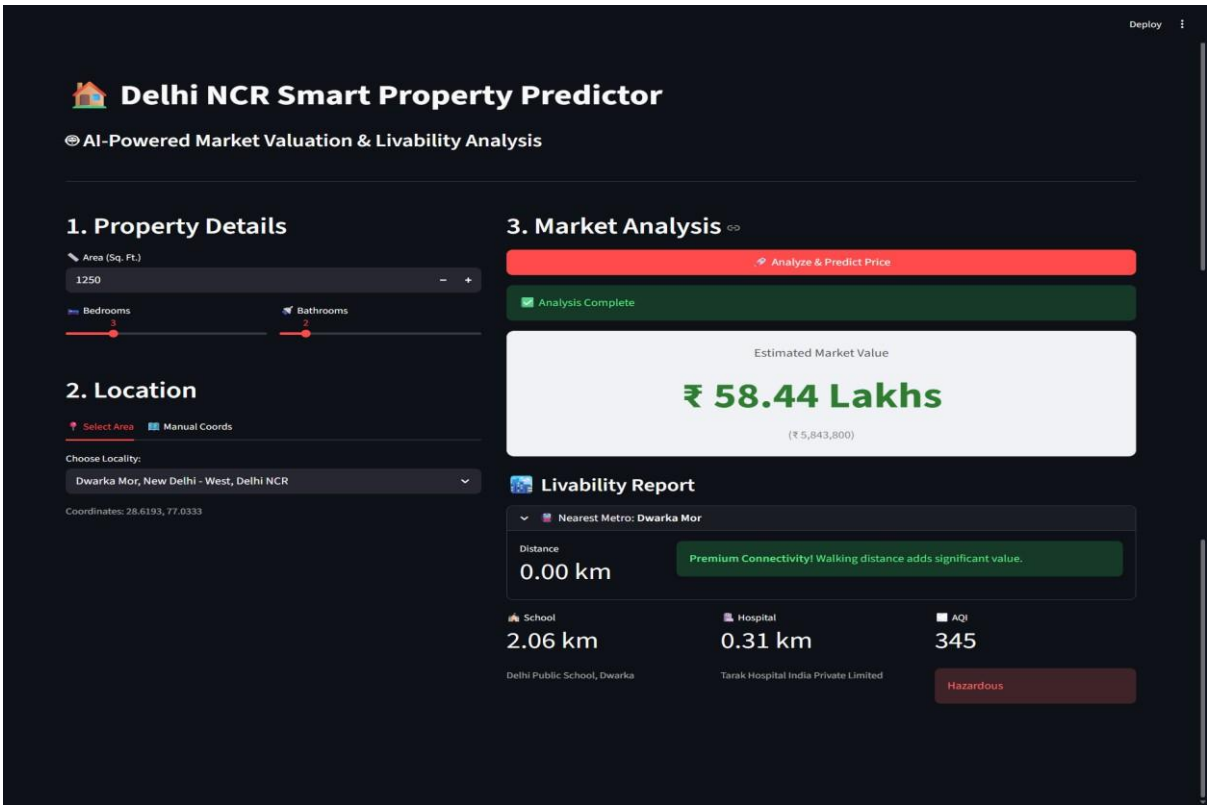
Geospatial Location (City/Sector): The city variable eg Gurgaon versus Ghaziabad serves as a price multiplier for properties located within those respective areas. A property located in Gurgaon has a base price that is almost 200% higher than that of an equivalent property located in Ghaziabad. This is due in part to the fact that the city of Gurgaon is considered to be a corporate hub.

Distance to Metro: The model showed that there is a significant negative correlation between price and Distance to Nearest Metro Station. Properties that are located within 0 – 1 km of the metro station are generally priced approximately 15% – 20% higher than properties that are located more than 3 km from a metro station. The metro premium is greatest for properties located on the Yellow Line (Gurgaon Metro) and Blue Line (Noida Metro).

6.3. The "Pollution Paradox" (Air Quality vs Price)

The relationship between air quality (AQI) and property prices across the globe is often seen as a contradiction. The rise in pollution-related health problems has resulted in decreased values of real estate assets, yet in Delhi, some of the most expensive micro-markets (e.g., South Delhi and Cyber City) have some of the worst AQI ratings because of a high volume of vehicle traffic.

The findings from recent market studies indicate that over the last five years, property values in Delhi have increased approximately 38% and in Gurgaon by approximately 84% despite declining air quality. Therefore, this indicates that economic potential and infrastructure capabilities are presently being valued higher than environmental health in the buyer's utility function. Pollution has yet to be properly accounted for when valuing properties within the metropolitan area, most likely due to the lack of available space for further building in desirable areas.



Price Prediction

7. Conclusion

7.1. Summary of Results

The current study successfully demonstrates that the Machine Learning methodology, applyingXGBoost as a technique, offers a robust and accurate framework for estimating residential property valuations in the National Capital Region (NCR) of Delhi. Furthermore, this research demonstrates the significant role that

location plays in property valuations, particularly in terms of connections to the Metro system and the associated economic characteristics of the surrounding neighbourhood.

The following are some key findings from this study:

The Value of Connectivity: The Delhi Metro has a significant influence on property values, with neighbourhoods around Metro stations being more valuable than neighbourhoods located further away from the Metro.

Arbitrage of Data: While traditional real estate valuation models fail to capture the economic signals present in the written descriptions of properties, this study demonstrates that our feature engineering capabilities enable us to identify these signals and include them in our models.

Market Resilience: The Delhi NCR property market is resilient to the effects of negative externalities, such as pollution, because of the demand for prime locations.

7.2. Implications for Stakeholders

Buyers can conclude from the model that there is a strong possibility of capital appreciation for newly Constructed Properties (under construction) located in the peripheral areas (e.g., Noida Extension, Aqua Line). However, it is critical to manage the execution risk associated with these properties.

Sellers/Developers should use connectivity (proximity to metro) and amenities (i.e., modular kitchens, park view) as primary factors when creating their listing descriptions because modern property valuation models will give a significantly higher importance to the connectivity and amenities found at a property than other property attributes.

Policymakers should consider using a value capture financing strategy when determining how to fund new Infrastructure projects. Value capture financing is a method of funding new infrastructure by coupling the increased value of the land surrounding the transit system with the increased revenue collected from taxes or other types of funding related to the transit system.

7.3. Future Directions

Future versions of this work have the opportunity to improve the ability of this research to predict housing prices by: Coming up with a model that uses temporal data based on transaction dates to map the timing of various stages of the housing cycle and inflation. Creating a model for using visual images of properties to assess their condition, such as through computer vision (CNN). Incorporating local environmental features such as noise levels and green space, in order to develop an improved "livability" index for areas.

References

1. R. Bamane, D. Gavhane, and P. Igave, "House Price Prediction Using Machine Learning," International Journal of Advanced Research in Science, Communication and Technology, vol. 5, no. 4, pp. 2110–2115, Oct. 2025.
2. S. Deepu, "Performance Evaluation Of Real Estate Price Predictor Using Machine Learning," SSRN Electronic Journal, 2024, doi: 10.2139/ssrn.4856312.
3. S. Fandian and M. Panwar, "Socio-economic impact of the Delhi Metro Line extension in Gurugram: A mixed-methods study on residential property prices," Journal of Contemporary Urban Affairs, vol. 9, no. 2,
4. pp. 1–15, 2025.
5. Y. Fu, "Comparison of Random Forest and Linear Regression for House Price Prediction," in Highlights in Science, Engineering and Technology, vol. 23, pp. 1–7, 2022.
6. K. K. Sahu, "Impact of metro railway network expansion on residential land prices in the National Capital Territory of Delhi," M.S. thesis, Graduate School of Public Policy, University of Tokyo, Tokyo, Japan, 2016.

7. M. Ruhela, F. Ahamad, S. Bhardwaj, and G. Singh, "Impact of Metro Construction Activities on Air Quality: A Case Study of Delhi Region," AgroEnvironmental Sustainability, vol. 2, no. 4, pp. 151–158, 2024.
8. J. Tapia, N. Chavez-Garzon, R. Pezoa, P. Suarez-Aldunate, and M. Pilleux, "Comparing automated valuation models for real estate assessment: Machine learning vs Hedonic pricing," PLoS ONE, vol. 20, no. 3, Art. no. e0318701, 2025.
9. S. S. Nair, A. Suresh, and M. L. Pai, "Comparative Analysis of Machine Learning Models for Real Estate Price Prediction in Urban City," in Proc. 2025 International Conference on Computing, Communication and Security (ICCCS), 2025, pp. 1–6.
10. R. Metcalfe and S. Roth, "The value of air quality in housing markets: Evidence from the Clean Air Act," Grantham Research Institute on Climate Change and the Environment, Working Paper 422, Feb. 2025.
11. R. W. Sinnott, "Virtues of the Haversine," Sky and Telescope, vol. 68, no. 2, p. 159, 1984.