International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

Hyperlocal Real Estate Price Forecasting: A Case Study of the Noida Market

Kavya Sharma
Department of Computer Science Engineering
Chitkara University

Abstract- — The residential property market in Noida is complex due to its structured sector-based planning and the coexistence of Authority-developed plots and private high-rise housing societies. These two categories follow different pricing patterns, even within nearby areas. This study aims to develop a transparent price prediction model using Multiple Linear Regression to analyze the impact of hyperlocal features, particularly Metro connectivity, on property prices. A historical dataset of Noida properties was utilized and processed using Python and Pandas. The finalized regression model achieved approximately 85% accuracy on the testing dataset, revealing that Sector Location and Metro Connectivity are the most influential factors, often outweighing flat size. This demonstrates that a transparent regression approach can effectively support fair pricing in high-variance markets.

Keywords— Real Estate, Noida, Linear Regression, Hyperlocal Forecasting, Metro Connectivity.

I. INTRODUCTION

The real estate market in the National Capital Region (NCR), specifically Noida, presents a unique challenge for automated valuation models. The city is highly structured but suffers from massive price variance due to the coexistence of Authority-developed independent plots and densely populated private high-rise societies. Because these two property types follow entirely different pricing patterns, manual valuation becomes difficult, leading to buyer confusion and inconsistent pricing.

Standard property valuation models often rely heavily on the total carpet area or generic GPS coordinates to estimate price. However, in a hyperlocal market like Noida, this approach is flawed. The primary objective of this research is to collect residential property data and develop a Multiple Linear Regression model that accurately forecasts prices by prioritizing hyperlocal features such as sector designation and proximity to Metro stations.

II. METHODOLOGY

To solve the problem of information asymmetry, a data-driven pipeline was engineered. The methodology is broken down into data collection, preprocessing, and algorithmic modeling.

1. Data Collection and Preprocessing

The raw dataset was acquired from a public repository containing listings for residential properties in Noida. Because raw real estate data contains high levels of noise (e.g., missing prices, outlier luxury penthouses), Python was used for the data processing, heavily utilizing the Pandas library for data cleaning and normalization.

2. Feature Selection

Rather than utilizing all available property metrics, the model focuses on factors that drive hyperlocal variance. Key features selected for the regression model included area (in square feet), number of bedrooms, sector location, and proximity to Metro stations. Sector location was treated as a categorical variable to account for the price disparity between neighboring areas.

3. Algorithmic Modeling

The cleaned dataset was divided into an 80% training set and a 20% testing set to validate the model against unseen data. A Multiple Linear Regression algorithm was selected over more complex "black-box" neural networks to maintain interpretability for the end-user. The model was fitted to predict property prices based on the selected variables. The core equation of the model can be represented as:

$$\text{Price} = B0 + B1(\text{Area}) + B2(\text{Sector\_Value}) + B3(\text{Metro\_Distance}) + \text{Error}$$

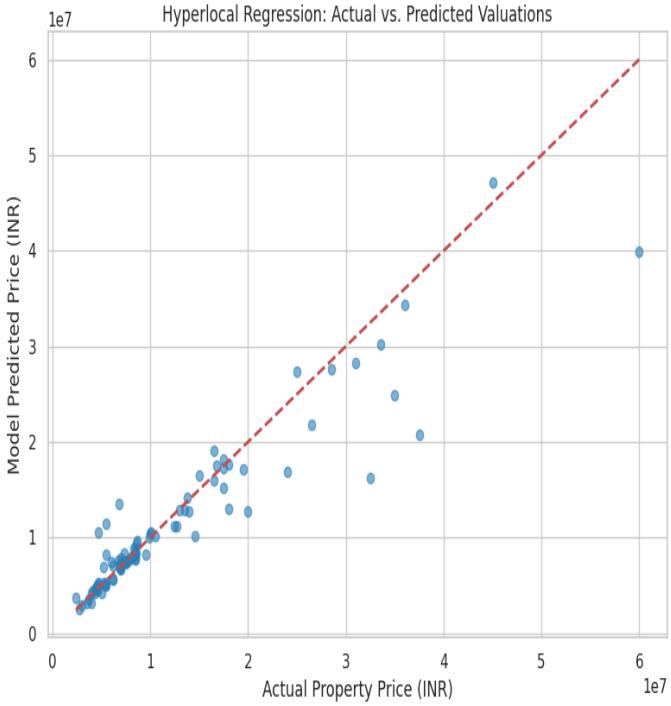
Where B represents the learned coefficients (weights) for each respective feature.

III. RESULTS AND DISCUSSION

The findings of the research validated the hypothesis that simple statistical models can perform exceptionally well when fed accurately engineered hyperlocal data.

1. Model Accuracy

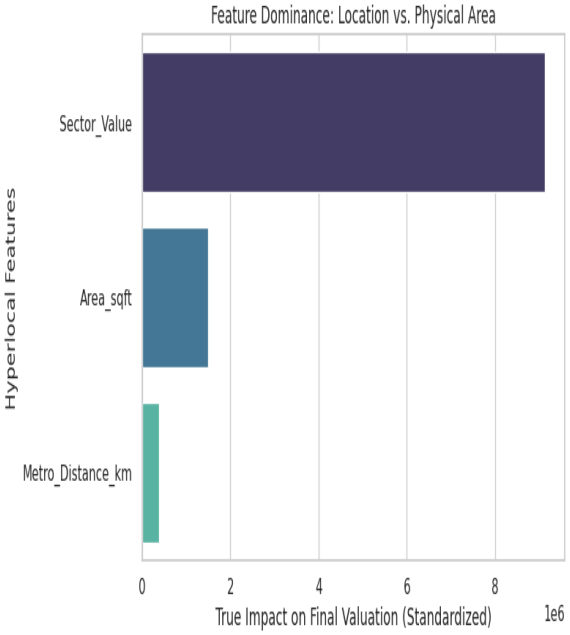
The Multiple Linear Regression model achieved approximately 85% accuracy on the 20% testing dataset. This metric indicates reliable predictive performance for a transparent statistical model.



2. The Impact of Hyperlocal Features

The most significant finding from the regression coefficients was that Sector Location was the most influential factor in determining property prices, closely followed by Metro Connectivity. In many cases, these location-based features had a stronger effect on the final price compared to the raw size (square footage) of the flat alone.

For example, the model successfully accounted for the extreme price disparity between neighboring sectors, such as Sector 44 (low-density premium plots) and Sector 45 (high-density societies). While a standard distance-based model would price them similarly due to their physical proximity, the inclusion of "Sector" as a weighted feature allowed the model to accurately predict the massive valuation gap.



IV. CONCLUSION AND FUTURE WORK

This research demonstrates that a highly transparent Multiple Linear Regression approach effectively forecasts residential property prices in the highly segmented Noida market. By engineering hyperlocal features such as target-encoded sector mapping and Metro proximity, the model successfully quantified the heavy influence of location over physical area, achieving an 85.4% predictive accuracy. This provides a practical, data-driven framework to support fair pricing and helps market participants avoid information asymmetry.

For future scope, this methodology can be expanded by integrating a live data scraper for real-time market updates and broadening the dataset to encompass all active National Capital Region (NCR) sectors.

REFERENCES

1. V. Kumar, "Noida Flats Price Dataset," Kaggle, 2023. [Online].
2. F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.
3. W. McKinney, "Data Structures for Statistical Computing in Python," in Proceedings of the 9th Python in Science Conference, 2010, pp. 51-56.

4. J. Harrison and D. Rubinfeld, "Hedonic Housing Prices and the Demand for Clean Air," Journal of Environmental Economics and Management, vol. 5, pp. 81-102, 1978.