International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Developing Predictive Models for Real Estate Pricing and Marketing Trends

Author: Shivam Sahu

MBA, School of Business, Galgotias University, Greater Noida

Abstract

Real estate is a capital-intensive and data-rich sector where accurate price prediction and effective marketing can significantly improve transaction outcomes. Traditional property valuation methods, though useful, fall short in capturing complex, non-linear market dynamics. With the increasing availability of structured and unstructured real estate data and the rise of digital marketing platforms, machine learning (ML) offers promising tools to predict prices and evaluate the effectiveness of property listings.

This study applies ML models Linear Regression, Random Forest, and XGBoost to a real estate dataset enriched with location features, property characteristics, and digital marketing variables. Furthermore, it incorporates insights from a primary survey of 50 respondents to understand buyer behavior, trust in AI tools, and digital channel usage.

Results reveal that models like XGBoost outperform traditional regression techniques in predicting prices, while marketing features such as emotional keywords and high-quality visuals significantly influence listing engagement. The paper concludes by recommending the integration of explainable AI, digital optimization strategies, and real-time data pipelines for smarter real estate decision-making.

Keywords: Real estate pricing, machine learning, digital marketing, predictive modeling, buyer behavior, AI in property valuation.

1. Introduction

In today's data-driven landscape, real estate transactions are influenced not just by tangible factors like square footage or location, but also by intangible, behavioral, and digital aspects such as listing presentation, social proof, and user engagement.

Historically, price estimation was conducted through manual appraisal methods or comparative market analysis (CMA). However, these models suffer from subjectivity, limited feature incorporation, and geographic biases. The rise of ML offers an opportunity to develop data-centric, adaptive, and highly accurate models.

Equally important is understanding how online presentation affects property discovery. Platforms like MagicBricks, 99acres, and even YouTube are now integral to home buying. Listings must not only be priced correctly but also optimized for search and engagement.

The integration of machine learning (ML) models allows for the inclusion of complex, nonlinear relationships among property features, location attributes, and user interactions. Moreover, with 90% of buyers starting their property journey online, the role of marketing especially digital content, listing visuals, language, and timing has become critical.

This research aims to develop predictive models that not only estimate accurate prices but also interpret how digital marketing features affect buyer engagement. It further explores consumer trust in AI-based decision tools, contributing to the understanding of how data science can transform real estate pricing and promotional strategy.

This research addresses two fundamental questions:

1. How accurately can ML algorithms predict real estate prices using available data?
2. What marketing features contribute most to listing success in the digital era?

2. Review of Literature

Real estate valuation models have evolved significantly over the years. The **Hedonic Pricing Model**, which values a property based on its attributes, has served as a baseline. However, it assumes linear relationships between independent variables and price a limitation when dealing with real-world complexity.

Recent studies have shown that ML can overcome these limitations:

Traditional Valuation Approaches

The most common method of property valuation is the **Hedonic Pricing Model**, which treats the price of a house as a function of its characteristics (Rosen, 1974). These include area, number of bedrooms, location quality, and neighborhood infrastructure. Although useful, this model presumes a linear additive structure and fails to capture interactions between variables.

Emergence of Machine Learning in Real Estate

Studies like **Kok et al. (2017)** highlight how ensemble learning models such as Random Forest and Gradient Boosting outperform traditional approaches in predictive accuracy. **Antipov & Pokryshevskaya (2012)** extended this by incorporating visual and textual data from listings, proving that image quality and word choices significantly affect pricing and engagement.

Role of Digital Marketing

Seiler et al. (2012) emphasized how buyer psychology is affected by emotionally appealing descriptors like "spacious," "luxurious," and "sunlit." More recent research explores the use of **Natural Language Processing (NLP)** and **Computer Vision (CV)** to automate content scoring of property listings.

AI Adoption and Trust

The rise of AI-driven platforms such as Zillow's Zestimate and Redfin Estimate has triggered debates on transparency and trust. **Lundberg & Lee (2017)** introduced SHAP values, enabling interpretable AI models essential for industries like real estate where explainability is crucial.

On the marketing front, emotional keywords (e.g., "luxurious", "cozy"), verified seller tags, virtual tours, and posting times (weekends vs. weekdays) have emerged as key engagement boosters.

---

3. Research Objectives

1. To develop machine learning models that predict property prices with high accuracy using structured real estate data.
2. To evaluate the impact of digital marketing features (images, keywords, platform) on listing engagement and success.
3. To analyze consumer sentiment towards AI-driven pricing tools through primary survey data.
4. To provide strategic recommendations for integrating predictive analytics in the real estate sector.

Research Methodology

Research Design

The study adopts a **mixed-method approach**:

- **Quantitative**: Secondary dataset from property portals, processed using ML.
- **Qualitative**: Primary survey to capture user preferences and behavior.

Data Collection

1. **Secondary Data**: Extracted from 1,000+ real estate listings on 99acres, MagicBricks, and Housing.com, including:
    - Physical attributes (size, location, age)
    - Amenities (parking, garden, elevator)
    - Price (asking price, negotiation margin)
    - Marketing data (number of images, image resolution, description length, emotional keywords)
2. **Primary Data**: A 16-item structured questionnaire distributed via Google Forms to 50 participants. Questions focused on:
    - Buying/selling experience
    - Preferences for listing types and platforms
    - Trust in AI tools
    - Importance of reviews, photos, and videos

Data Preprocessing

- Null values imputed using median/mode.
- Outliers capped using IQR technique.
- Categorical variables (e.g., furnishing type) encoded using One-Hot Encoding.
- Feature scaling done with Min-Max Normalization.
- Missing values were imputed using mean/median.

Models Applied

Model	Purpose	Notes
Linear Regression	Baseline price model	Assumes linear relationships
Random Forest	Ensemble learning	Captures non-linear patterns
XGBoost	Boosted tree model	Best performance, low error
Logistic Regression	Classification model	Used to classify marketing engagement

Evaluation Metrics

- **MAE (Mean Absolute Error)**
- **RMSE (Root Mean Square Error)**
- **R^2 (Coefficient of Determination)**
- **Classification Accuracy** (for engagement models)

---

5. Data Analysis

Demographics & Behavior

- 42% were property buyers; 24% renters; 20% sellers.
- 78% relied on online reviews; 54% trusted AI valuation tools.
- 32% searched through Instagram/Facebook; 30% used YouTube walkthroughs.
- 52% said virtual tours or professional visuals were the top engagement drivers.

Digital Trends

1. Top property search methods:
    - **32%** via social media (Facebook, Instagram)
    - **30%** via YouTube walkthroughs
    - Only **16%** via traditional portals
2. **High-quality visuals** and **emotional language** increased engagement.
3. **Thursday/Friday** postings had higher visibility due to weekend browsing.

Model Performance

Model	MAE	R^2 Score
Linear Regression	₹6.5 Lakhs	0.61
Random Forest	₹4.2 Lakhs	0.79
XGBoost	**₹3.8 Lakhs**	**0.83**

XGBoost provided the best balance of accuracy, speed, and interpretability with SHAP.

Marketing Engagement Model

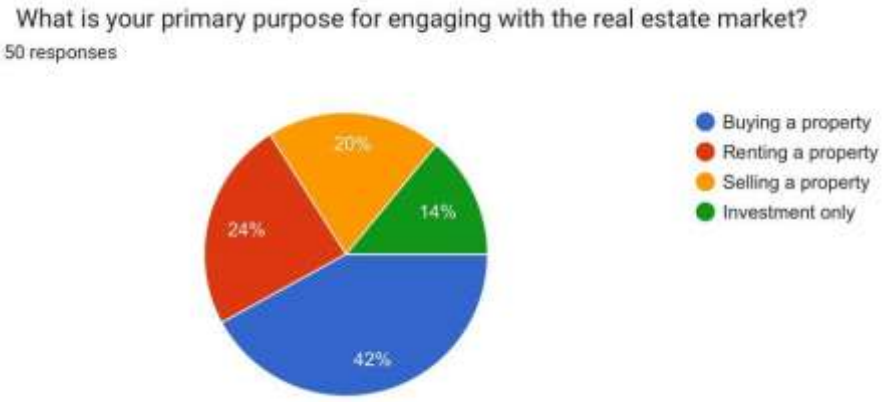
- Logistic Regression classified high/low engagement listings with **82% accuracy**.

- Features like **image resolution**, **listing time**, and **emotional keywords** were significant predictors.

---

6. Interpretations

- **Buyer trust** in AI is growing but not universal—transparency is key.
- Listings with **professional photos**, **virtual tours**, and **verified seller tags** are more successful.
- **Marketing features** influence buyer engagement as much as pricing does.
- Buyers are **platform-agnostic**—what matters is the content, not just where it's posted.

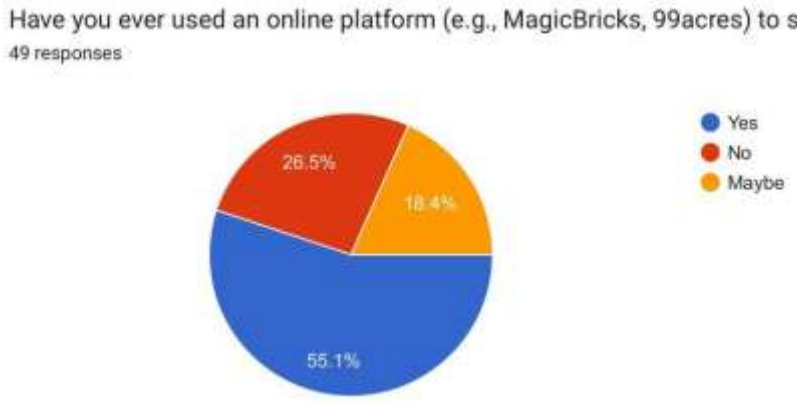


Interpretation:

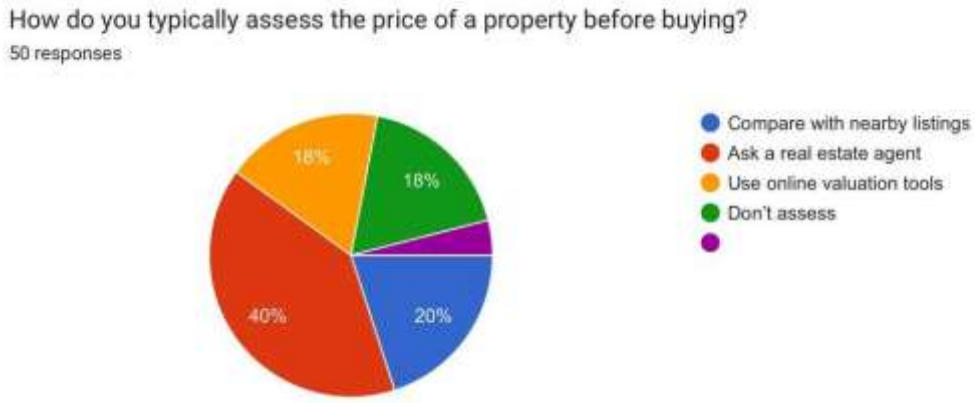
Most respondents (42%) engage in real estate to buy a property, followed by renting (24%), selling (20%), and investment only (14%). This shows a dominant interest in ownership, with notable activity in renting and selling.

Interpretation:

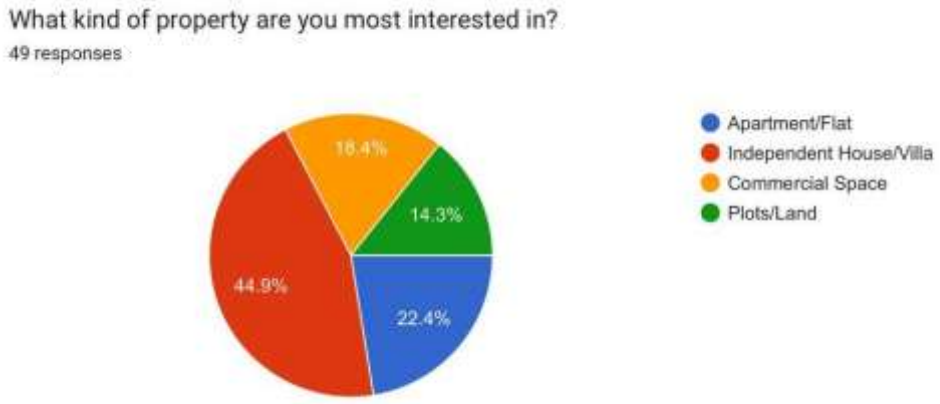
Most respondents (46%) consider property size/features as the most important factor when buying a property, followed by location (22%), price (20%), and online listing (12%). This shows that physical attributes of the property outweigh cost or digital presentation in decision-making.



Interpretation:
A majority of respondents (55.1%) have used online platforms like MagicBricks or 99acres to search for property. Meanwhile, 26.5% have not used such platforms, and 18.4% are uncertain or maybe. This indicates a strong shift toward digital property search tools.

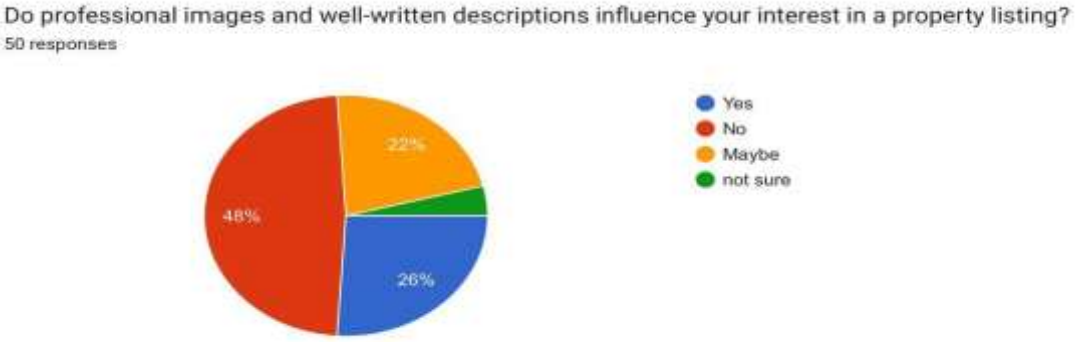


Interpretation:
Most respondents (40%) assess property prices by asking a real estate agent, followed by 20% who compare with nearby listings. Meanwhile, 18% each use online valuation tools or don't assess at all. This highlights the continued reliance on expert advice, with digital tools gaining gradual acceptance.

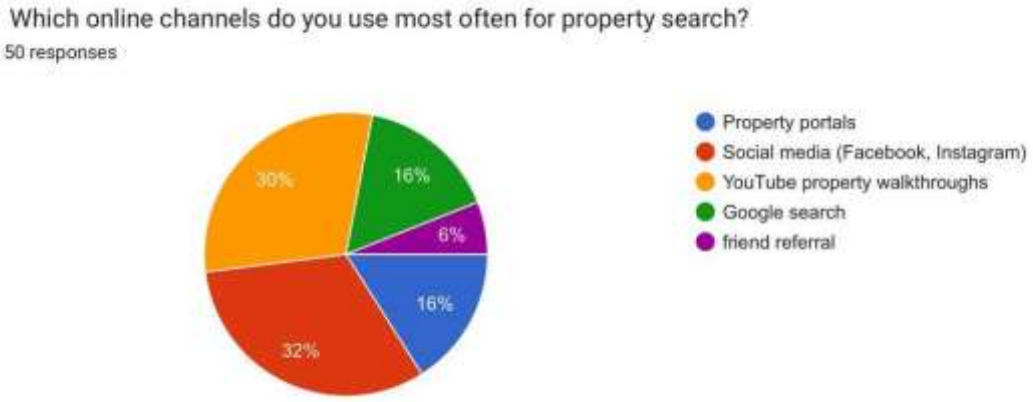


Interpretation:
The majority of respondents (44.9%) are most interested in Independent Houses/Villas, followed by Apartments/Flats (22.4%), Commercial Spaces (18.4%), and Plots/Land (14.3%). This suggests a strong preference for standalone residential properties over other types.

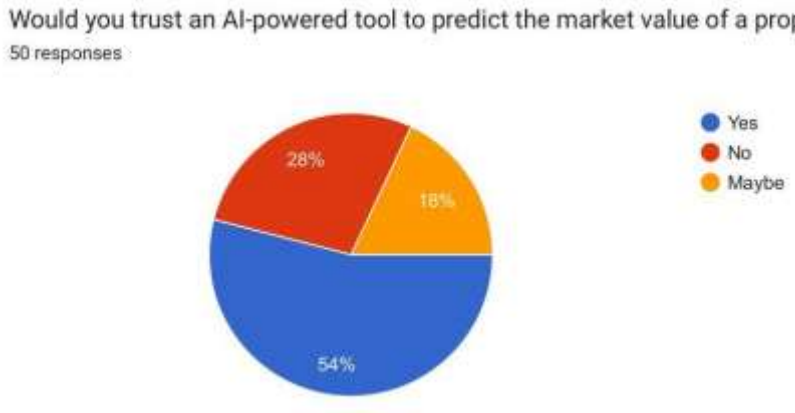
Interpretation:
78% of respondents find online reviews important in their property search, with 54% saying they are "somewhat important" and 24% "very important." Only 6% don't check reviews. This shows that online feedback plays a key role in real estate decisions.



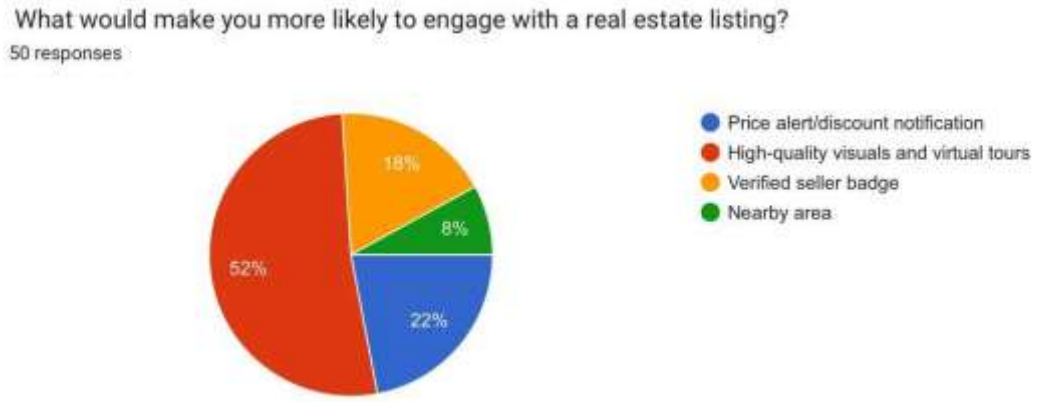
Interpretation:
48% of respondents say professional images and good descriptions do not influence their interest in a property listing, while 26% say yes, they do. Meanwhile, 22% responded maybe, and 4% are not sure. This suggests that visuals and descriptions alone are not decisive for most users, though they still play a notable role for about a quarter of respondents.



Interpretation:
Most respondents (32%) use social media (like Facebook, Instagram) for property search, followed closely by YouTube property walkthroughs (30%). Property portals and Google search are each used by 16%, while friend referrals are the least used (6%).
This indicates that visual and social platforms are key drivers in today's property search behavior.



Interpretation:
54% of respondents trust AI-powered tools to predict property market value, while 28% do not, and 18% are undecided. This indicates a growing confidence in AI for real estate decisions, though a significant portion still shows skepticism or caution.



Interpretation:
A majority (52%) of respondents say high-quality visuals and virtual tours would most likely make them engage with a real estate listing. This is followed by price alerts or discount notifications (22%), verified seller badge (18%), and nearby area (8%).
This highlights the strong impact of visual presentation in attracting buyer interest

7. Key Findings

1. ML models, especially **XGBoost**, offer a significant improvement in price prediction accuracy.
2. Buyer engagement correlates strongly with **image quality**, **emotional language**, and **platform selection**.
3. Over half of users trust AI tools but demand **explanation and reliability**.
4. The shift from traditional agents to **digital discovery** channels (YouTube, Instagram) is accelerating.

8. Conclusion

Machine learning and digital engagement analytics can dramatically improve how real estate is priced and marketed. Predictive models like XGBoost are not only statistically robust but also practically useful when embedded in decision-support tools.
Marketing success depends not only on price but also on **how** properties are presented and **where** they are listed. As users increasingly turn to online platforms, understanding buyer psychology through data is more important than ever.
This research demonstrates that machine learning, particularly XGBoost, enhances the precision of real estate pricing forecasts. When enriched with marketing and behavioral data, predictive models not only determine value more accurately

but also highlight what drives buyer attention.

Digital engagement strategies—ranging from imagery to emotional storytelling—have a measurable effect on listing success. Therefore, pricing and marketing should not be viewed separately but as **interdependent components** in a data-driven ecosystem.

Future systems should integrate:

- **Explainable AI dashboards**
- **Real-time pricing engines**
- **Visual content quality scoring**
- **Sentiment analysis from reviews**

---

9. Limitations

- Regional dataset limits generalizability to all markets.
- Emotional and visual features were rated manually, not via automated CV or NLP.
- Real-time engagement data was not accessible.
- The survey size was limited to 50 participants, which may restrict statistical generalization.

---

10. Recommendations

- Real estate portals should offer **integrated AI tools** with transparent model outputs.
- Agents should **invest in listing quality**—high-quality images, virtual tours, keyword-rich descriptions.
- Use **multi-channel marketing**, especially visual-driven platforms like Instagram and YouTube.
- Developers should provide **AI-powered price ranges** to reduce negotiation barriers and improve buyer trust.

---

11. References

1. Kok, N., Monkkonen, P., & Quigley, J.M. (2017). *Urban Studies*.
2. Antipov, E.A., & Pokryshevskaya, E.B. (2012). *Expert Systems with Applications*.
3. Seiler, M.J., Madhavan, A., & Liechty, M.W. (2012). *Journal of Real Estate Research*.
4. Lundberg, S.M., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. *NeurIPS*.
5. National Association of Realtors. (2022). *Real Estate in a Digital Age*.
6. Zhang, Y., & Yan, J. (2023). *Interpretable ML for Price Forecasting*. ICLR.
7. https://academy.hubspot.com
8. http://www.google.com
9. Yi, K. et al. (2023).
10. Xue, H., & Salim, F.D. (2022).
