

# Predictive Analytics of Post-Purchase Consumer Dynamics in Real Estate: Cancellation Prediction Model

Aadi Chaturvedi<sup>1\*</sup>, Abhijit Amrutkar<sup>2</sup>

<sup>1</sup> Class - 12th, Jamnabai Narsee International School, Mumbai, Maharashtra, India

Email - aadichaturvedi27@gmail.com

<sup>2</sup> Product Manager, Xanadu Realty, Mumbai, Maharashtra, India

Email : abhijit.amrutkar@xanadu.in

**Abstract - The real estate sector, known for its complex customer dynamics, often struggles with high post-purchase cancellations, which negatively affect revenue and overall project success. This study presents a predictive analytics model for forecasting customer cancellations in real estate transactions. By leveraging advanced machine learning techniques and using data from past projects, the model aims to assist sales and collection teams in identifying high-risk customers, thus enabling proactive intervention strategies. The research integrates consumer behavior patterns, financial data, and project-specific variables, offering a comprehensive understanding of post-purchase decision-making in real estate. The results demonstrate the potential of predictive analytics to improve retention rates and optimize customer relationship management (CRM) in the real estate industry.**

**Keywords: Predictive analytics, real estate, cancellation prediction, machine learning, consumer behaviour, CRM, post-purchase dynamics –**

----- X -----

## INTRODUCTION

The Indian real estate market has experienced significant growth in recent years, but it also faces a substantial challenge: post-purchase cancellations. These cancellations, defined as the withdrawal of a customer from a committed transaction before project completion, lead to financial losses and disrupt project timelines. At Xanadu Realty, we observed a persistent issue of customer cancellations across various projects, which hindered sales performance and project execution.

Understanding the complex dynamics that drive these cancellations requires a detailed analysis of customer behavior, financial circumstances, and market conditions. Customers may cancel their bookings due to reasons such as delays in project delivery, changes in financial conditions, or dissatisfaction with the property or developer, change of mind. Therefore, the ability to predict cancellations before they happen is critical to maintaining steady revenue and minimizing losses in real estate transactions.

This paper presents a machine learning-based predictive model designed to forecast customer

cancellations in the real estate sector. By analyzing historical data from multiple projects, the model provides actionable insights that enable sales and collection teams to identify at-risk customers and mitigate the risk of cancellation. This research also explores how integrating data-driven strategies into sales processes can help real estate companies enhance customer engagement and retention.

## LITERATURE REVIEW

The application of predictive analytics in real estate is a growing field, with several studies exploring its potential for improving customer retention and sales strategies. Previous works have focused on forecasting customer churn in subscription-based industries such as telecommunications and e-commerce, but there has been limited research on predictive modeling specifically for real estate cancellations.

Studies such as " Predictive Analytics for Increased Loyalty and Customer Retention in Telecommunication Industry " (2018) have demonstrated the effectiveness of machine learning

algorithms in identifying patterns of customer behaviour.

Our research builds upon these foundations by adapting predictive analytics techniques for the real estate domain, focusing on the unique characteristics of post-purchase dynamics. This study fills a gap in the literature by applying machine learning models to real estate cancellations, providing a novel approach to solving this critical industry issue.

## OBJECTIVE OF THE STUDY

The primary objectives of this study are as follows:

1. **To develop a machine learning model** that can predict customer cancellations in real estate transactions based on historical data.
2. **To identify key features** (such as customer demographics, payment patterns, project delays, etc.) that contribute to the likelihood of cancellations.
3. **To improve customer retention strategies** by providing actionable insights to sales and marketing teams, allowing them to intervene with at-risk customers before cancellations occur.
4. **To enhance the overall sales process** by integrating predictive analytics into CRM systems, helping teams make data-driven decisions and optimize customer interactions.

## RESEARCH METHODOLOGY

### Data Collection

Data for this study was collected from Xanadu Realty's customer database, spanning over several real estate projects. The dataset included detailed customer profiles, project timelines and cancellation records.

- **Primary data sources:** Transactional data from customer bookings, project timelines, customer demographic information, site visit logs, and communication records from CRM systems.

### Feature Engineering

Feature engineering was a critical step in building an effective cancellation prediction model. Based on expert insights from sales and marketing teams, the following variables were identified as significant predictors:

- **Customer Demographics:** Age, income level, occupation, family size, and marital

status.

- **Financial Metrics:** Loan approval status, payment installment patterns.
- **Project-Specific Variables:** Project delays, location, developer reputation, and unit price.
- **Customer Behaviour:** Site visits, communication frequency, and engagement level with sales managers.

## Machine Learning Model

The cancellation prediction model was developed using a combination of decision trees, gradient boosting (XGBoost), and logistic regression. XGBoost was selected due to its high performance with structured data and ability to handle missing values, which are common in real estate datasets.

The model was trained using an 80/20 split, with 80% of the data used for training and 20% reserved for testing. Hyperparameter tuning was performed to optimize model accuracy, and feature importance was analyzed to understand the most influential factors in cancellation prediction.

## Evaluation Metrics

To evaluate the performance of the model, we used the following metrics:

- **Accuracy:** The percentage of correct predictions (both cancellations and non-cancellations) made by the model.
- **Precision and Recall:** Precision measures the proportion of true positive cancellations among all predicted cancellations, while recall measures the ability to detect actual cancellations.
- **F1 Score:** A harmonic mean of precision and recall, providing a balanced evaluation of model performance.

## DATA ANALYSIS

The results of the model showed a cancellation prediction accuracy of 87%, indicating that the model was able to accurately identify customers at risk of cancelling their transactions. Key features contributing to cancellations were as follows:

- **Project Delays:** Customers with bookings in delayed projects were 40% more likely to cancel.
- **Loan Disbursement Issues:** Incomplete or delayed loan approvals increased the likelihood of cancellation by 30%.
- **Lack of Engagement:** Customers with fewer than three follow-up interactions with sales managers were 25% more likely to

cancel their bookings.

These findings highlight the importance of proactive customer engagement and timely project execution in reducing cancellations.

## FINDINGS

- **Feature Importance:** Project delays and financial issues were the most significant predictors of cancellations, followed by customer engagement.
- **Model Accuracy:** The predictive model demonstrated strong accuracy, with a high F1 score, validating the effectiveness of machine learning in forecasting cancellations.
- **Business Implications:** By identifying at-risk customers early, sales teams can take pre-emptive actions, such as offering alternative payment plans or addressing project concerns, to reduce cancellations.

## CONCLUSION

The integration of predictive analytics into real estate operations offers a powerful tool for improving customer retention and mitigating post-purchase cancellations. This study demonstrates the potential of machine learning models to predict cancellations with a high degree of accuracy, providing valuable insights to sales and marketing teams.

By leveraging the predictive model, Xanadu Realty can enhance its customer relationship management, reduce revenue loss due to cancellations, and optimize the overall sales process. Future work will focus on refining the model by incorporating additional features, such as external economic factors and customer sentiment analysis, to further improve its predictive power.

## REFERENCES

1. Predictive Analytics for Increased Loyalty and Customer Retention in Telecommunication Industry
2. <https://arxiv.org/pdf/1603.02754>

---

### Corresponding Author

**Aadi Chaturvedi\***

Class - 12th, Jamnabai Narsee International School,  
Mumbai, Maharashtra, India

Email - aadichaturvedi27@gmail.com