© APR 2026 | IRE Journals | Volume 9 Issue 10 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I10-1716183

# Predictive Customer Lifetime Value: Transitioning from Segment-Based Classification to Probabilistic Revenue Forecasting

DEVANSH MISHRA[1], DEEPAM SINGH[2], DR. ISHRAT ALI[3], SANJAY PACHAURI[4]
[1,2,3,4]*Department of Computer Science Engineering, Greater Noida Institute of Technology*

***Abstract- Earlier approaches to Customer Lifetime Value (CLV) have largely focused on segmentation techniques such as RFM analysis combined with clustering methods like K-Means [1]. While these methods are useful for identifying customer groups, they do not directly estimate how much value a customer will generate in the future. In this paper, we propose a predictive framework that moves beyond descriptive segmentation toward revenue forecasting. The approach combines the BG/NBD probabilistic model to estimate whether a customer is still active [2] with an XGBoost regression model to predict future spending [3]. This hybrid setup allows businesses to move from static customer grouping to a more dynamic and practical forecasting system.***

## I. INTRODUCTION

Customer Relationship Management (CRM) is gradually shifting from analyzing past behavior to predicting future actions. Traditional models based on the Recency, Frequency, and Monetary (RFM) metrics has been widely used to group customers into segments [1]. These methods are the simple and effective, but they are mainly describe what has already happened .

A key limitation of these methods is that their inability to distinguish between the customer churn and temporary inactivity. This creates uncertainty, which can hinder the development of effective retention strategies .

To address this, the current work introduces a second-phase framework that focuses on prediction rather than description. The goal is not just to label customers, but to estimate their future monetary contribution more precisely.

## II. METHODOLOGICAL TRANSITION: FROM V1 TO V2

The proposed framework builds upon earlier work and extends it into a predictive setting. In the first phase, clustering and classification were used to identify customer segments.

In the second phase, the focus shifts from forecasting revenue at an individual level .

| Feature | Phase 1 | Phase 2 |
|---|---|---|
| Objective | Customer Segmentation | Revenue Prediction |
| Approach | Unsupervised | Hybrid (Probabilistic + ML) |
| Model | K-Means + XGBoost Classifier | BG/NBD + XGBoost Regressor |
| Output | Segment Labels | Continuous Value |
| Metrics | Precision, Recall, F1 | RMSE, MAE |

This shift makes the system more useful in real-world applications, where knowing how much a customer will spend matters more than just knowing their segment.

## III. FEATURE ENGINEERING

To improve model performance, additional features were introduced along with standard RFM metrics [1].

Inter-Purchase Time (IPT): Helps in understanding how frequently a customer makes purchases.

Monetary Volatility: Captures how consistent or inconsistent a customer's spending pattern is.

© APR 2026 | IRE Journals | Volume 9 Issue 10 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I10-1716183

Tenure: Indicates how long the customer has been associated with the business.

These features provide more depth and allow the model to better capture customer behavioral patterns.

## IV. MODELING APPROACH

4.1 BG/NBD for Customer "Aliveness"
One of the key challenges in CLV prediction is identifying whether , a customer is still active state. The BG/NBD model helps to estimate the probability that a customer is "alive" based on past transaction data [2].

Instead of assuming churn directly, this method treats it as a probabilistic outcome, which is more realistic in most business scenarios.

4.2 XGBoost for Revenue Prediction
To estimate future revenue, an XGBoost regression model is used [3].

Unlike classification models, this directly predicts a continuous monetary value.

The model takes engineered features along with BG/NBD outputs as input. RMSE is used as main evaluation metric to ensure that the prediction errors are minimized, especially for high-value customers .

## V. EXPLAINABILITY AND BUSINESS USE

For the model to be more useful in practice, it is important to understand its predictions. SHAP values are used to explain how different features influence the output [4].

This makes it easier to identify the pattern such as declining engagement or inconsistent spending. Based on these insights, businesses can take actions like offering discounts or targeting specific customers for retention campaign.

## VI. CONCLUSION

This paper presents shift from the traditional segmentation methods to a predictive framework for customer lifetime value. By combining this probabilistic modeling with machine learning, the approach provides more accurate and actionable insights .

Overall, this proposed method helps businesses to better understand their customers and make informed decisions about marketing and retention strate5gies.

## REFERENCES

[1] Bult, J. R., & Wansbeek, T. (1995). "Optimal Selection for Direct Mail." Marketing Science, 14(4), 378–394.

[2] Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005). "Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model." Marketing Science, 24(2), 275–284.

[3] Chen, T., & Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[4] Lundberg, S. M., & Lee, S.-I. (2017). "A Unified Approach to Interpreting Model Predictions." Advances in Neural Information Processing Systems (NeurIPS).