Global Journal of Engineering and Technology Advances
eISSN: 2582-5003
Cross Ref DOI: 10.30574/gjeta
Journal homepage: https://gjeta.com/

(Review Article)

Check for updates

Machine learning ensemble techniques for customer lifetime value prediction in e-commerce businesses: A review

Amorue Daniel *, Ayankoya F. Y and Kuyoro S. O

Department of Computer Science, Babcock University.

Global Journal of Engineering and Technology Advances, 2026, 27(03), 124-128

Publication history: Received on 25 April 2026; revised on 08 June 2026; accepted on 10 June 2026

Article DOI: https://doi.org/10.30574/gjeta.2026.27.3.0145

Abstract

Online retailers increasingly depend on precise Customer Lifetime Value (CLTV) estimation for data-driven Customer Relationship Management. Traditional frameworks, including RFM heuristics, probabilistic Buy-Till-You-Die models, and Markov chain architectures, offer interpretability but fail when confronted with highly skewed, zero-inflated behavioral datasets common in digital commerce. This review synthesizes the transition from legacy methods to ensemble learning techniques, focusing on homogeneous bagging (Random Forests), sequential boosting (GBM, XGBoost, CatBoost), and heterogeneous stacked generalization. The study identifies critical success factors, performance trends across asymmetric optimization targets, and dataset variations. This study serves as a definitive reference to guide future CLTV deployment pipelines and validation rigor.

Keywords: Customer Lifetime Value; Ensemble learning; CLTV prediction; Bagging; Boosting; Stacking

1. Introduction

The ubiquity of the internet and internet enabled devices has encouraged the adoption of online shopping by customers around the world. As more global businesses enter into the online shopping scene, they need efficient techniques for offering customer centric solutions for sustaining existing customers. The access to abundant customer data and predictive analytic solutions provide opportunities for businesses to offer targeted customer experiences that optimize the customer engagement and spending. Customer lifetime value (CLTV) is a business metric that enables organizations to analyze and estimate the financial profit their customers bring to business over a period of time. Precision in CLTV prediction enables e-commerce businesses to target customers with campaigns proportional to their contributions. This improves customer experiences and retention while optimizing business resource management [1],[2].

According to global statistical organization Statista [3] 2026 revenue in the e-commerce market is projected to hit $3.88trillion US dollars. This showcases vast opportunities and a highly competitive market space with a large number of businesses looking to optimize their profits while delivering quality customer service. E-commerce organizations keep investing huge amounts of marketing capital to attract and sustain relationships with profitable customers [5], but developing a successful Customer Relationship Management Strategy, also involves knowing every customer's value to the business.

Marketing efforts, socio-economic conditions and dynamic customer behaviors, such as irregular purchase frequencies leading to highly skewed data distributions [4], [5] make CLTV prediction a daunting task for e-commerce businesses. According to [6], businesses invest about 5 to 10 times more resources when acquiring new customers than to retain them, this fact encourages business to engage in retaining valuable customers considering they are the backbone of business.

* Corresponding author: Amorue Daniel

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0.

This review addresses the fragmentation in CLTV prediction literature, and provides an opportunity to synthesize and observe the most efficient techniques presently available. The scope of this review is based on available peer reviewed literature predicting customer lifetime value for businesses. This review analyzes traditional probabilistic methods and advanced machine learning ensemble techniques, providing a definitive methodological reference to guide future CLTV prediction in enterprise and academic environments.

2. Traditional CLTV prediction approaches

Traditional methods have always been used in the estimation of customer lifetime value before the intervention of machine learning techniques, one of the most popular methods is the RFM methodology introduced by Blattberg and Van der Scheer 1995. The RFM model utilizes customer characteristics including recency of last purchase, frequency of the customer purchase, and monetary value contributed to the business over a set period. [7] explored the use of RFM as a technique for developing a customer selection and marketing resource allocation framework, and it showed that customers chosen based on their lifetime value offered higher profits over time when they were compared to customers selected based on other metrics. The limitation of the RFM model is its reliance on only the recency, frequency and monetary values, and overlooks other dataset features such as temporal values, marketing campaigns, and customer demography.

Pareto-NBD and BG-NBD are popular Buy-Till-You-Die (BTYD) probabilistic models that are typically used to predict customer lifetime value, in continuous and non-contractual business environments, where the customers can drop off without notifying the business. [8] explored the application of the RFM analysis for CLTV estimation using the Pareto/NBD model in an e-commerce environment, showing the efficiency of CLTV in the online retail business. A major draw of the BTYD CLTV prediction models is they expect customer purchase behaviors to remain consistent over time and they often fail to account for other dynamic factors that influence the customers.

The Markov Chain Model is an advanced stochastic model that can be used for estimating customer lifetime value, by modeling how customers move between different states such as active, churned, loyal, or varying spending tiers. [9] developed a novel hybrid model integrating machine learning and Markov Chain Modelling, to predict lifetime value in a Business-to-Business environment. The framework reportedly modelled customer state transitions based on product and service score significantly enhancing predictive accuracy while providing realistic predictions for future customer value. Markov chain expresses that the likelihood of customers states changing to a specific state relies on the current state. It acknowledges the dynamic nature of customer spending behavior over time instead of treating them as a static property, thereby adding a level of efficiency to them.

3. Machine learning ensemble techniques

There are three main categories of ensemble machine learning algorithms: bagging, boosting, and stacking.

3.1. Bagging

Bagging also known as bootstrap aggregation [10], is an ensemble learning method that combines multiple instances of a machine learning model that has been trained on randomly selected subsets of a main dataset. These random data subsets are selected with replacement through the bootstrapping technique. The models are then trained in parallel and an average of their final results (in regression tasks) or a majority vote (in classification tasks) is performed to derive the final result. Bagging reduces variance and bias, and this is great for predictive analysis in business intelligence environments where customer behavior usually shows high variability. Working with the diverse data subsets it can easily capture and understand the diversity in customer behavior across the datasets. A prevalent example of bagging is Random Forest [11] which is an ensemble of several Decision trees. [12] developed a framework for predicting customer retention and profitability, the study showed that random forest when deployed consistently outperformed the ordinary linear regression, highlighting its efficiency.

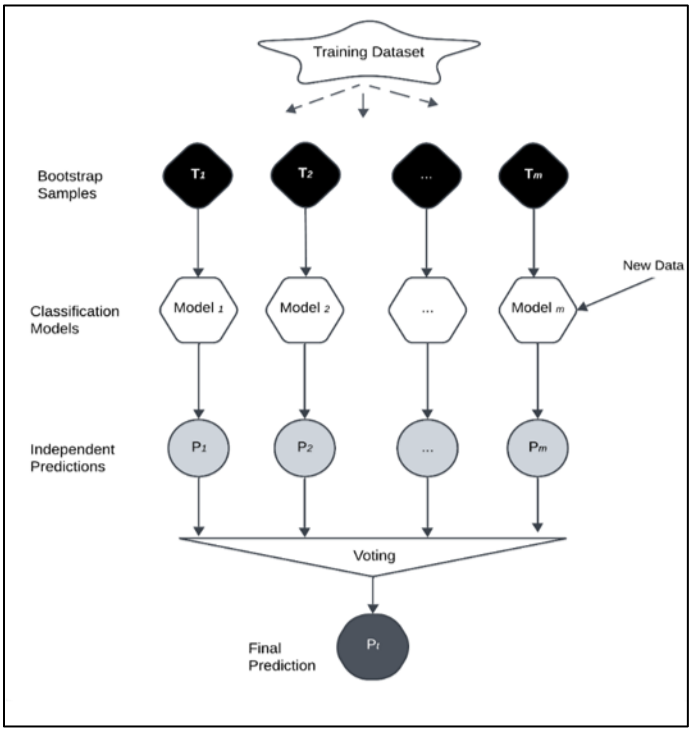

Figure 1 Structure of Bagging Technique

3.2. Boosting Technique

Boosting is an ensemble learning technique that trains multiple weak machine learners in a sequence to create a strong predictive model. It trains models in a sequential and iterative process so the succeeding learner focuses on correcting the errors of the previous learners, thereby improving model performance. During the training process, misclassified samples are given higher weights so the succeeding learners focus more on them. This process is repeated until a satisfactory level is reached for stopping, the final result is a weighted result of the base learners, creating a stronger model with lower variance and bias [14]. [15] developed a model using gradient boosting machines (GBM) in prediction of CLTV across multiple industries including ecommerce, telecommunications, and financial services, evaluating the result against existing models such as linear regression models and random forests. The GBM model consistently outperformed the other models, underscoring its potential as an efficient tool for CLTV prediction across multiple industries. Boosting techniques are particularly useful in predictive analytics for business intelligence majorly because of their ability to handle large and multi-dimensional datasets typically found in modern business environments. Some examples of boosting techniques are AdaBoost, XGBoost, Gradient Boosting Machines, and Cat Boost.

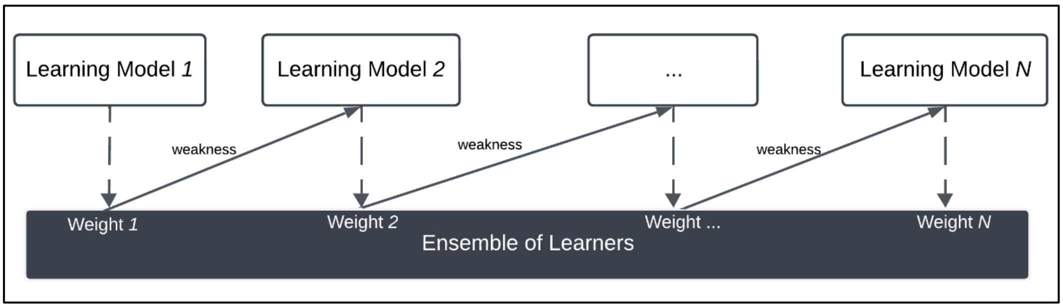

Figure 2 Structure of Boosting Technique

3.3. Stacking

Also known as stacked generalization the stacking technique works by putting together multiple base learners on a first layer and then aggregating their output into a single final model known as the meta learner. The base models are trained independently before their various outputs are combined to serve as input for the final learner. The base learners are usually diverse allowing the ensemble model to benefit from the strengths of all its learners. A stack ensemble is usually as good as its strongest base learner, the base model can be a combination of random forest, XGBoost, and LightGBM which later feed their outputs into a Linear Regression meta-learner. [17] proposed a novel hybrid model for improving CLTV prediction using machine learning, by combining Elastic Net, Random Forest, XGBoost, SVM (as base learners) and Linear Regression (as the meta learner). It utilized the diversity of learners improving predictive accuracy.

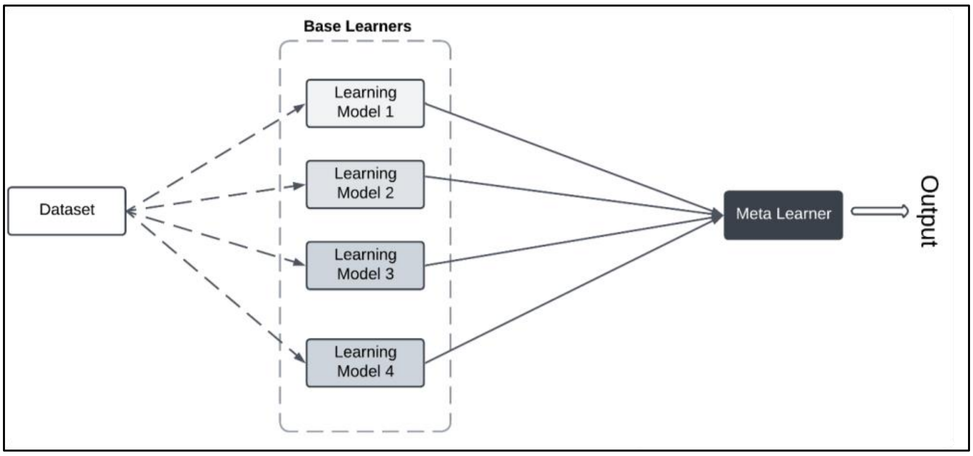

Figure 3 Structure of Stacking Technique

4. Conclusion

This review synthesized the transition from traditional CLTV prediction methods which included RFM heuristics, Pareto/NBD and BG/NBD probabilistic models, and Markov chain state-transition architectures, to advanced ensemble learning techniques. Traditional approaches offer high structural interpretability and remain valuable for baseline predictions. However, they consistently fail when confronted with the highly skewed, zero-inflated, and multi-modal behavioral datasets that characterize modern e-commerce environments.

Ensemble learning addresses these fundamental limitations through three distinct mechanisms. Bagging (Random Forest) reduces variance by training multiple models in parallel on bootstrap samples, making it particularly effective for noisy customer data while preserving interpretability through feature importance scores. Boosting (GBM, XGBoost, CatBoost) trains models sequentially, with each subsequent learner focusing on correcting errors of previous models. This sequential error-correction mechanism proves especially valuable for sparse customer purchase data where many customers make few transactions. Stacking combines diverse base learners often mixing random forest, XGBoost, and support vector machines with a meta-learner that learns optimal weighting of base model predictions.

Our synthesis reveals three key findings. Stacking achieves the highest predictive accuracy for CLTV but requires significant computational resources (approximately 5-10x slower than single models). Boosting offers the optimal balance of accuracy and efficiency for most e-commerce applications, particularly when customer purchase data is sparse. Bagging remains the preferred choice when stakeholders require interpretable feature importance explanations for marketing decisions.

This review serves as a definitive methodological reference to guide algorithmic selection, deployment pipelines, and validation rigor for CLTV prediction in enterprise and academic analytics environments.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] V. Kumar and B. Rajan, "Social coupons as a marketing strategy: a multifaceted perspective," *J. Acad. Mark. Sci.*, vol. 40, no. 1, pp. 120–136, Jan. 2012, doi: 10.1007/s11747-011-0283-0.

[2] S. Gupta and D. R. Lehmann, "Customers as assets," *J. Interact. Mark.*, vol. 17, no. 1, pp. 9–24, 2003.

[3] "eCommerce - Worldwide | Statista Market Forecast." Accessed: Jun. 03, 2026. [Online]. Available: https://www.statista.com/outlook/emo/ecommerce/worldwide?srsltid=AfmBOooyT3v41KFRBmKMDHzPIqrFypBH875OaVIzFzkEpLhD117kQqnj#revenue

[4] E. C. Malthouse and R. C. Blattberg, "Can we predict customer lifetime value?," *J. Interact. Mark.*, vol. 19, no. 1, pp. 2–16, 2005.

[5] D. C. Schmittlein, D. G. Morrison, R. Colombo, S. Gupta, D. R. Lehmann, and D. J. Poirier, "Counting Your Customers: Who-Are They and What Will They Do Next?," *Manage. Sci.*, vol. 33, no. 1, pp. 1–24, Jan. 1987, doi: 10.1287/mnsc.33.1.1.

[6] A. Gallo, "The value of keeping the right customers. Harvard Business Review," 2014, *October*.

[7] R. Venkatesan and V. Kumar, "A customer lifetime value framework for customer selection and resource allocation strategy," *J. Mark.*, vol. 68, no. 4, pp. 106–125, 2004.

[8] R. A. Megantara *et al.*, "RFM Analysis for Customer Lifetime Value with PARETO/NBD Model in Online Retail Dataset," *Monet. J. Keuang. dan Perbank.*, vol. 11, no. 2, pp. 284–290, Oct. 2023, doi: 10.32832/moneter.v11i2.409.

[9] S. Nasiri, "Enhanced Business-to-Business Lifetime Value Prediction Using a Hybrid Machine Learning and Markov Chain Model," *J. Business-to-bus. Mark.*, pp. 1–16, Oct. 2025, doi: 10.1080/1051712X.2025.2574597.

[10] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, pp. 123–140, 1996.

[11] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.

[12] B. LARIVIERE and D. VANDENPOEL, "Predicting customer retention and profitability by using random forests and regression forests techniques," *Expert Syst. Appl.*, vol. 29, no. 2, pp. 472–484, Aug. 2005, doi: 10.1016/j.eswa.2005.04.043.

[13] S. Raschka, Y. H. Liu, and V. Mirjalili, *Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python*. Packt Publishing Ltd, 2022.

[14] E. Ding, "Ensemble (Boosting, Bagging, and Stacking) in Machine Learning: Easy Explanation for Data Scientists," Ensemble (Boosting, Bagging, and Stacking) in Machine Learning: Easy Explanation for Data Scientists. Accessed: Nov. 10, 2024. [Online]. Available: https://www.youtube.com/watch?v=sN5ZcJLDMaE&t=309s

[15] [15] T. T. Bukhari, O. Oladimeji, E. D. Etim, and J. O. Ajayi, "Customer lifetime value prediction using gradient boosting machines," *Gyanshauryam, Int. Sci. Ref. Res. J.*, vol. 5, no. 4, pp. 488–506, 2022.

[16] Z. Zhang, "Boosting Algorithms Explained," Towards Data Science. Accessed: Dec. 10, 2024. [Online]. Available: https://towardsdatascience.com/boosting-algorithms-explained-d38f56ef3f30

[17] A. M. Haddadi and H. Hamidi, "A hybrid model for improving customer lifetime value prediction using stacking ensemble learning algorithm," *Comput. Hum. Behav. Reports*, vol. 18, p. 100616, May 2025, doi: 10.1016/j.chbr.2025.100616.

[18] Yash, "Ensemble Stacking for Machine Learning and Deep Learning," analytics vidhya. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/08/ensemble-stacking-for-machine-learning-and-deep-learning/