Predicting Future High-Value Customers in E-Commerce: An Explainable Machine Learning Framework for Digital Customer Relationship Management

Yaxin Chi

Chinese Agriculture University, No.17, QinghuaEastRoad, HaidianDistrict, Beijing, 100083, China

bbw131507@163.com

Abstract. Digital customer relationship management increasingly depends on converting transaction histories into actionable forward-looking signals. This study develops a temporally separated and explainable machine learning framework for identifying customers who will enter the top quintile of future spending. Using the public UCI Online Retail dataset, 15 behavioral features are constructed from a 274-day observation window ending on 31 August 2011, and labels are determined from the following 100-day prediction window. Records with missing customer identifiers, cancellations or returns, and non-positive prices are removed, while customers with only one valid historical purchase are retained. Logistic regression, random forest, and XGBoost are evaluated by stratified five-fold cross-validation and by rolling-origin tests. The random forest achieves mean ROC-AUC of 0.861 and PR-AUC of 0.692. Its ROC-AUC remains 0.896, 0.861, and 0.846 under top-10%, top-20%, and top-25% definitions, respectively, and reaches 0.861 and 0.852 in two out-of-time tests. SHAP analysis identifies cumulative monetary value, product diversity, recent revenue, purchase frequency, and average basket value as the main predictive signals. The results support a transparent customer-prioritization workflow whose conclusions are not dependent on a single value threshold or a random cross-validation split.

Keywords: digital economy; e-commerce; customer value prediction; explainable machine learning; customer relationship management

1 Introduction

E-commerce firms accumulate large volumes of transaction data, but the managerial value of these data depends on whether past behavior can be converted into timely decisions. In digital customer relationship management (CRM), one recurrent problem is resource prioritization: firms must decide which existing customers should receive retention attention, personalized communication, or service resources before future value becomes observable. Historical sales totals are useful, but they are backward-looking.

© The Author(s) 2026
V. Erokhin et al. (eds.), *Proceedings of the 6th International Conference on Internet Finance and Digital Economy (ICIFDE 2026)*, Advances in Economics, Business and Management Research 412,
https://doi.org/10.2991/978-94-6239-830-6_16

A customer who generated high revenue in the past may become inactive, whereas a customer with moderate cumulative spending but accelerating purchase activity may become economically important in the next period. The digital-economy challenge is therefore not merely to describe customers, but to extract forward-looking signals from transaction streams.

Traditional customer analytics often summarize purchasing histories through recency, frequency, and monetary value (RFM). The RFM paradigm remains attractive because it is simple and behaviorally meaningful, and prior work has linked these measures to customer lifetime value (CLV) [2,3]. Recent studies continue to use RFM and engineered transaction features as foundations for customer valuation, segmentation, and future-value prediction [11,12,18]. However, segmentation and retrospective scoring do not directly answer a prediction question: given only information available today, which customers are most likely to become high-value customers during a future window? This distinction is important because a model evaluated on contemporaneous customer value may benefit from information that would not have been available when a real marketing decision was made.

Machine learning offers flexible tools for this prediction task. Tree-based ensembles can capture non-linear interactions among purchase frequency, transaction size, engagement duration, product breadth, and recent spending dynamics [4,5]. Recent CLV research has examined multi-output neural networks, interpretable stacked ensembles, and scalable retail prediction systems [9,12,13]. Yet greater predictive complexity creates a second problem for CRM: managers need to understand why customers are prioritized. A probability score without an explanation may be difficult to translate into differentiated actions and may encourage indiscriminate targeting. Explainable machine learning, particularly feature-attribution methods such as SHAP, provides a way to connect predictive performance with interpretable behavioral drivers [6,9,16,17].

This paper investigates whether historical online retail transactions can predict future high-value customer status using a compact and explainable computational framework. The study makes four contributions. First, it adopts a strict temporal design in which all predictors are generated before a fixed cutoff and labels are generated only from subsequent purchases. Second, it documents customer eligibility, return and cancellation handling, time-window lengths, outlier treatment, and all 15 features in a reproducible feature table. Third, it tests whether conclusions change when the high-value group is defined as the top 10%, top 20%, or top 25% of future spend. Fourth, it supplements random cross-validation with rolling-origin evaluation that trains on earlier snapshots and tests on later snapshots. The objective is not to claim a universally superior algorithm, but to identify a reproducible workflow suitable for digital CRM decisions in transaction-rich e-commerce settings.

2 Related Work and Research Gap

Customer value analysis has long emphasized the predictive content of observed purchasing behavior. Fader et al. [3] formally connected RFM characteristics with customer valuation, while Chen et al. [2] demonstrated how RFM-based data mining can

support customer-centric business intelligence in online retail. Recent peer-reviewed work has moved beyond static segmentation toward flexible prediction systems. Multi-output neural networks with SHAP have been used for e-commerce CLV prediction [9]; hierarchical and stacked ensembles have addressed heterogeneous customers, temporal constraints, and model complexity [10,12]; transaction-feature pipelines have supported noncontractual customer value analysis [11]; and scalable retail systems have emphasized deployment and cross-brand generalization [13]. Recent reviews confirm the growing role of business analytics and AI in CLV management while highlighting data quality, interpretability, and temporal validity as continuing concerns [14,18].

For the present research problem, three gaps remain operationally relevant. First, descriptive segmentation is different from forecasting a future value class, and a decision-support system should reproduce the information set available before the decision date. Second, the practical meaning of a high-value label depends on campaign capacity and should not be assumed to be invariant to a single percentile cutoff. Third, strong performance under random folds does not by itself demonstrate robustness when purchasing behavior changes over time. Cohort-based CLV evidence shows that behavioral shifts can materially alter out-of-sample prediction and the identification of top-value customers [15], while recent explainable customer analytics research emphasizes that predictions should be translated into actionable and auditable decisions [16,17]. This study therefore combines temporal separation, threshold sensitivity, rolling-origin testing, and model explanation in a compact customer-level classification framework.

3 Data and Method

3.1 Dataset, Preprocessing, and Temporal Prediction Design

The analysis uses the public Online Retail dataset from the UCI Machine Learning Repository [1]. The source contains 541,909 line-item records for a United Kingdom-based non-store retailer between 1 December 2010 and 9 December 2011. Each record includes invoice number, product code, quantity, invoice time, unit price, customer identifier, and country. Because the prediction unit is an identifiable purchasing customer, 135,080 records without a customer identifier are removed. Invoice numbers beginning with 'C', non-positive quantities, and non-positive unit prices are treated as cancellations, returns, or non-realized transactions and excluded. After these rules, 397,884 valid positive-purchase lines remain. Returned items are not netted against their original purchase; the target is realized positive spending in the future window.

A fixed cutoff of 1 September 2011 separates feature construction from outcome measurement. The observation window is the 274-day interval [1 December 2010, 1 September 2011), and the prediction window is the 100-day interval [1 September 2011, 10 December 2011), which includes transactions through 9 December. The observation window contains 226,467 valid lines and yields 3,317 eligible customers. Eligibility requires at least one valid purchase before the cutoff; no minimum number of invoices or minimum tenure is imposed. Consequently, customers with short histories remain in the sample. Customers with no valid purchase in the prediction window are

retained and assigned future spending of zero. The prediction window contains 171,417 valid lines.

For customer i, future spending S_i^future is the sum of quantity multiplied by unit price in the prediction window. The main outcome identifies customers at or above the 80th percentile of the eligible-customer distribution:

$$y_i(q) = 1\{S^{future}_i \geq Qq(S^{future})\}, \quad q \in \{0.75, 0.80, 0.90\}.$$

The top-20% definition is selected as the primary specification because a quintile is a common, easily communicated segmentation rule and creates a customer pool that is selective but still large enough for retention and service campaigns. In this dataset it corresponds to future spending of at least GBP 961.26 and 664 customers (20.02%; the slight difference from exactly 20% is caused by ties). It is an operational resource-allocation segment, not a universal economic definition of value. To assess dependence on this choice, the full evaluation is repeated for top-10% and top-25% labels. A genuine long-horizon CLV label cannot be observed because only 100 days remain after the cutoff and the dataset lacks margin, acquisition cost, and churn information; future spending is therefore described as a short-horizon realized-value proxy rather than lifetime value.

3.2 Behavioral Feature Engineering and Transformations

Fifteen customer-level predictors are derived exclusively from valid transactions before each cutoff. Table 1 provides the complete definitions, formulas, input windows, and transformations. Revenue is quantity multiplied by unit price. Invoice-based quantities are calculated after aggregating line items to the invoice level, preventing larger baskets from mechanically receiving greater weight in weekend share. For customers without purchases in a recent block, recent revenue and frequency are set to zero. Tenure is lower-bounded at one day, and the +1 terms in the momentum ratio make the feature defined when either 60-day block has zero revenue.

Table 1. Complete behavioral feature specification. W = winsorization using the 0.5th and 99.5th percentiles estimated on the training fold; L = log(1+x); S = standardization using training-fold mean and standard deviation.

Feature	Definition/formula	Input window	Logistic regression	Tree models
Recency	(cutoff - last valid purchase)/day	All history	W + S	W
Frequency	Number of distinct valid invoices, F	All history	W + L + S	W
Monetary value	Σ quantity × unit price, M	All history	W + L + S	W
Average basket value	M/F	All history	W + L + S	W
Average items per invoice	Σ quantity/F	All history	W + L + S	W
Product diversity	Number of distinct StockCode values	All history	W + L + S	W
Active months	Number of distinct calendar months with a valid purchase	All history	W + L + S	W
Tenure	max[1, (cutoff - first valid purchase)/day]	All history	W + L + S	W
Recent revenue	Revenue in [cutoff-60 days, cutoff), R60	Final 60 days	W + L + S	W

Feature	Definition/formula	Input window	Logistic regression	Tree models
Recent frequency	Distinct invoices in [cutoff-60 days, cutoff)	Final 60 days	W + L + S	W
Recent revenue share	R60/M	All history + final 60 days	W + S	W
Spending momentum	(1+R60)/(1+Rprev60)	Final and preceding 60-day blocks	W + L + S	W
Purchase rate	30F/tenure	All history	W + L + S	W
Weekend share	Weekend invoices/F	All history	W + S	W
UK indicator	1 if modal customer country is United Kingdom; 0 otherwise	All history	S	Unchanged

Extreme customer-level spending and count values are not removed, because unusually large wholesale purchases are part of the retailer's economic reality. Instead, every continuous predictor is winsorized within each training fold and the learned limits are applied unchanged to the corresponding validation or test data. Logistic regression additionally applies log(1+x) to right-skewed spending, count, tenure, momentum, and rate variables and standardizes the inputs. Recency, shares, and the binary country indicator are standardized without a log transform. Random forest and XGBoost use the winsorized values directly. Fitting all transformations only on training data prevents information from validation or future snapshots from influencing preprocessing.

3.3 Models, Evaluation, and Explanation

Three classifiers are compared: class-weighted logistic regression, random forest [4], and XGBoost [5]. Logistic regression uses L2 regularization with C = 1 and a balanced class weight. The random forest uses 200 trees, maximum depth 5, square-root feature sampling, and a balanced class weight. XGBoost uses 200 trees, maximum depth 3, learning rate 0.05, subsample and column-sample ratios of 0.8, and a negative-to-positive class ratio as the positive-class weight. All stochastic procedures use random seed 42.

The main performance estimates use stratified five-fold cross-validation. The study reports ROC-AUC, precision-recall AUC (PR-AUC), F1, and balanced accuracy. PR-AUC is emphasized because the positive class is deliberately smaller than the negative class and precision-recall analysis is informative for imbalanced classification [7]. F1 and balanced accuracy use the default probability threshold of 0.5. A separate stratified 70/30 split is used only for ROC visualization and global explanation. The strongest cross-validated model is explained by mean absolute SHAP values on the holdout customers [6,16]. All data processing and modeling are implemented in Python using pandas, NumPy, scikit-learn [8], XGBoost, and SHAP.

3.4 Threshold Sensitivity and Rolling-origin Validation

Threshold sensitivity is evaluated by repeating the five-fold random-forest analysis after redefining the positive class as the top 10% (90th percentile) and top 25% (75th percentile) of future spending. This tests whether model discrimination is specific to

the primary top-quintile segment and shows how precision-recall and F1 change when campaign capacity is narrower or broader.

Temporal robustness is evaluated with three customer snapshots. The first snapshot uses transactions before 1 March 2011 and labels spending in [1 March, 1 June); the second uses transactions before 1 June and labels spending in [1 June, 1 September); the third is the main 1 September snapshot and labels spending in [1 September, 10 December). The random forest is first trained on the March snapshot and tested on the June snapshot. It is then trained on the combined March and June snapshots and tested on September. The 80th-percentile threshold is recomputed within each future window (GBP 784.00, GBP 629.77, and GBP 961.26, respectively). Each customer-snapshot is treated as a row, customer identifiers are never model inputs, and all transformations are re-estimated from the earlier training snapshots only. The future label window for every training origin ends before the next test origin, so later purchasing behavior cannot leak into earlier training records.

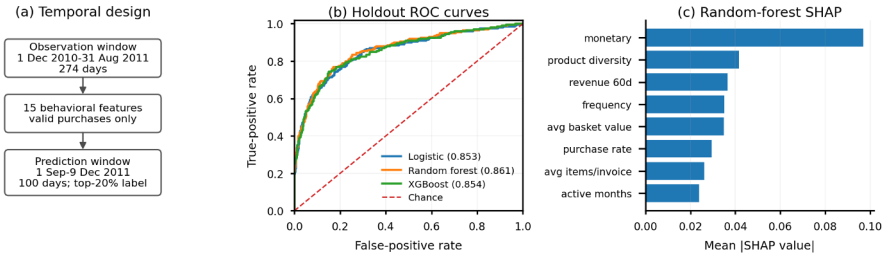


Fig. 1. Temporal prediction design and model diagnostics. (a) A 274-day observation window produces 15 features for a separate 100-day future-spending label. (b) Holdout ROC curves for the three classifiers. (c) Mean absolute SHAP values for the eight most influential random-forest features.

4 Results

Table 2 shows that the three classifiers achieve broadly similar discrimination. The random forest records the highest mean ROC-AUC (0.861) and PR-AUC (0.692), while XGBoost has the highest mean balanced accuracy by a negligible margin (0.778 versus 0.777 for the random forest). The random forest F1 score is 0.607, compared with 0.605 for XGBoost and 0.589 for logistic regression. The close results indicate that the engineered behavior features contain substantial separable signal and that additional model complexity does not create a large performance gain.

The independent holdout curves in Fig. 1(b) are directionally consistent with cross-validation. Holdout ROC-AUC is 0.861 for the random forest, 0.854 for XGBoost, and 0.853 for logistic regression. The corresponding holdout PR-AUC values are 0.712, 0.702, and 0.700. These differences are modest, supporting a conservative interpretation: ensemble learning provides a small improvement, but a well-specified linear benchmark remains competitive.

Table 2. Study sample and five-fold cross-validation performance. Values are mean ± standard deviation across folds.

Panel A. Temporal sample design				
Full source dataset	541,909 line-item records			
Valid positive-purchase data	397,884 line-item records			
Observation window	1 Dec 2010-31 Aug 2011 (274 days; 226,467 records)			
Prediction window	1 Sep-9 Dec 2011 (100 days; 171,417 records)			
Eligible customer sample	3,317 customers; ≥1 valid historical purchase			
Primary high-value definition	Future spend ≥ GBP 961.26 (top 20%)			
Positive class	664 customers (20.02%)			
Panel B. Five-fold cross-validation performance				
Model	**ROC-AUC**	**PR-AUC**	**F1**	**Balanced accuracy**
Logistic regression	0.857 ± 0.017	0.690 ± 0.035	0.589 ± 0.021	0.779 ± 0.015
Random forest	0.861 ± 0.010	0.692 ± 0.025	0.607 ± 0.024	0.777 ± 0.022
XGBoost	0.852 ± 0.017	0.686 ± 0.034	0.605 ± 0.022	0.778 ± 0.020

Table 3. Robustness analysis for the random forest. Panel A varies the customer-value cutoff under the main September origin; Panel B evaluates later customer snapshots using models trained only on earlier origins.

Panel A. Sensitivity to the high-value segment definition						
Definition	**Spend cutoff**	**Positive n (%)**	**ROC-AUC**	**PR-AUC**	**F1**	**Balanced accuracy**
Top 10%	GBP 1,639.58	332 (10.01%)	0.896 ± 0.026	0.631 ± 0.059	0.533 ± 0.047	0.804 ± 0.035
Top 20%	GBP 961.26	664 (20.02%)	0.861 ± 0.010	0.692 ± 0.025	0.607 ± 0.024	0.777 ± 0.022
Top 25%	GBP 748.17	830 (25.02%)	0.846 ± 0.024	0.714 ± 0.031	0.634 ± 0.035	0.765 ± 0.027
Panel B. Rolling-origin out-of-time validation						
Training origin(s)	**Test origin**	**Train/Test n**	**ROC-AUC**	**PR-AUC**	**F1**	**Balanced accuracy**
Mar 2011	Jun 2011	1,682/2,718	0.861	0.663	0.596	0.778
Mar + Jun 2011	Sep 2011	4,400/3,317	0.852	0.665	0.579	0.774

Panel A of Table 3 addresses the choice of a top-quintile label. The random forest remains discriminative across all three operational segments: ROC-AUC is 0.896 for the top 10%, 0.861 for the top 20%, and 0.846 for the top 25%. PR-AUC is not directly comparable across label definitions because its baseline rises with positive-class prevalence. As expected, the narrow top-10% campaign has lower F1 (0.533) but higher balanced accuracy (0.804), whereas the broader top-25% campaign has higher PR-AUC (0.714) and F1 (0.634). Thus, the conclusion that past behavior predicts future value does not depend on exactly selecting the 80th percentile, although the appropriate operating threshold should reflect campaign capacity and error costs.

Panel B of Table 3 provides a stricter temporal test. A model trained on the March snapshot achieves ROC-AUC 0.861 and PR-AUC 0.663 when applied to the June snapshot. Training on the March and June snapshots and testing on September yields ROC-AUC 0.852 and PR-AUC 0.665. F1 declines from 0.596 to 0.579, which is consistent with some change in the purchase distribution, but discrimination remains close to the random-fold result. The rolling-origin evidence therefore supports temporal

generalization while also showing why periodic recalibration is preferable to assuming a stationary customer process.

SHAP attribution provides a behavioral interpretation of random-forest predictions. Historical cumulative monetary value is the dominant feature, with mean absolute SHAP value 0.097. Product diversity is second (0.041), followed by revenue in the most recent 60 days (0.037), invoice frequency (0.035), average basket value (0.035), purchase rate (0.029), average items per invoice (0.026), and active months (0.024). The ranking combines value intensity, relationship breadth, and recent activity. A customer can therefore receive a high score because of consistently large spend, broader product engagement, strong recent revenue, repeated purchasing, or a high purchase rate rather than a single behavioral dimension.

5 Discussion

The findings have four implications for digital customer relationship management. First, the top-20% threshold should be understood as a decision-capacity choice rather than an intrinsic boundary between valuable and non-valuable customers. A firm with limited premium-service capacity may use the top 10%, while a lower-cost communication program may use the top 25%. Recent cohort-based CLV research likewise evaluates top-10% and top-20% customer groups and shows that the identification of best-value customers can vary across behavioral periods [15]. The sensitivity results show that ranking performance remains useful across these choices, but F1, precision-recall performance, and the monetary cutoff change materially. Thresholds should therefore be recalibrated to available resources, customer economics, and the cost of false positives and false negatives.

Second, transaction data can support a genuinely forward-looking customer-prioritization task when the modeling design respects time. Separating the observation and prediction windows is more aligned with real decisions than clustering customers using all available transactions. This concern is consistent with recent CLV studies that explicitly address temporal constraints, daily or repeated scoring, and cohort-specific behavioral change [10,13,15]. The rolling-origin results strengthen this evidence by showing that a model learned from earlier snapshots remains discriminative on later purchasing periods. The modest decline in F1 at the final origin also indicates that periodic retraining and drift monitoring are necessary in operational deployment.

Third, careful feature engineering may be at least as important as selecting a more complex algorithm. Random forest provides the strongest overall discrimination, but logistic regression and XGBoost remain close. Recent studies similarly show that useful CLV systems can be built with engineered behavioral variables, stacked ensembles, or hybrid pipelines rather than relying exclusively on highly complex architectures [9,11-14,18]. For small and medium-sized digital retailers, a compact behavioral feature layer can therefore be a practical intermediate architecture between spreadsheet-based RFM scoring and high-cost representation learning. The complete feature specification also makes the model easier to audit, reproduce, and maintain.

Fourth, model explanation can improve the managerial usability of predictive scores. The SHAP ranking translates model output into familiar transaction concepts: cumulative value, product breadth, recent revenue, frequency, basket size, and purchase rate. Recent customer analytics studies likewise use SHAP and counterfactual explanations to connect predictive outputs with understandable retention and customer-management actions [9,16,17]. These drivers can support differentiated actions. High-basket customers may justify premium service or product bundles; frequent customers may be suitable for loyalty interventions; customers with strong recent revenue may benefit from timely cross-selling. The model should not be treated as an autonomous rule for excluding lower-ranked customers. Predicted value is probabilistic, and firms should apply privacy safeguards, monitor drift, and retain standard service for all customers.

Several limitations define the scope of the evidence. The dataset represents one UK-based non-store retailer and only one year of transactions. Future spending over 100 days is not equivalent to CLV, and the absence of margin, acquisition cost, returns-linked-to-original-orders, campaign exposure, browsing behavior, and churn information prevents a full economic value model. Recent reviews similarly emphasize that CLV applications depend on the available value definition, business context, risk information, and analytical design [14,18]. Rolling-origin snapshots can contain repeated customers, which reflects periodic CRM scoring but means that the temporal tests assess distributional generalization rather than generalization to entirely unseen identities. The analysis is predictive rather than causal; high SHAP importance does not imply that changing a feature will create customer value. Future research should validate the framework across retailers, use longer horizons for true CLV estimation, compare probability ranking with profit-sensitive objectives, and combine prediction with uplift or causal models that estimate responsiveness to intervention.

6 Conclusion

This study presents a compact explainable machine learning framework for predicting future high-value customers from e-commerce transactions. By constructing all predictors in a clearly defined 274-day observation window and defining customer value only from a subsequent 100-day period, the design converts transaction histories into a realistic forward-looking classification task. Across 3,317 eligible customers, the random forest achieves mean ROC-AUC of 0.861 and PR-AUC of 0.692. Its discrimination remains useful under top-10%, top-20%, and top-25% value definitions and in two rolling-origin tests. SHAP analysis identifies cumulative monetary value, product diversity, recent revenue, purchase frequency, and average basket value as the principal predictive signals. The main practical implication is that transparent preprocessing, explicit time windows, threshold sensitivity, temporal validation, and model explanation can provide a reproducible basis for CRM prioritization without requiring an opaque high-complexity model.

Data availability

The Online Retail dataset is publicly available from the UCI Machine Learning Repository under CC BY 4.0 [1].

Code availability

The Python workflow used for cleaning, feature construction, within-fold transformations, threshold sensitivity, rolling-origin validation, model comparison, and SHAP explanation can be archived with the final conference submission.

References

1. D. Chen, "Online Retail," UCI Machine Learning Repository, 2015, doi: 10.24432/C5BW33.
2. D. Chen, S. L. Sain, and K. Guo, "Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining," Journal of Database Marketing & Customer Strategy Management, vol. 19, no. 3, pp. 197-208, 2012, doi: 10.1057/dbm.2012.17.
3. P. S. Fader, B. G. S. Hardie, and K. L. Lee, "RFM and CLV: Using iso-value curves for customer base analysis," Journal of Marketing Research, vol. 42, no. 4, pp. 415-430, 2005, doi: 10.1509/jmkr.2005.42.4.415.
4. L. Breiman, "Random forests," Machine Learning, vol. 45, pp. 5-32, 2001, doi: 10.1023/A:1010933404324.
5. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785-794, doi: 10.1145/2939672.2939785.
6. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems 30, 2017, pp. 4765-4774.
7. T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," PLOS ONE, vol. 10, no. 3, e0118432, 2015, doi: 10.1371/journal.pone.0118432.
8. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.
9. G. Yılmaz Benk, B. Badur, and S. Mardikyan, "A new 360° framework to predict customer lifetime value for multi-category e-commerce companies using a multi-output deep neural network and explainable artificial intelligence," Information, vol. 13, no. 8, art. 373, 2022, doi: 10.3390/info13080373.
10. S. Curiskis, X. Dong, F. Jiang, and M. Scarr, "A novel approach to predicting customer lifetime value in B2B SaaS companies," Journal of Marketing Analytics, vol. 11, no. 4, pp. 587-601, 2023, doi: 10.1057/s41270-023-00234-6.
11. Y. Sun, H. Liu, and Y. Gao, "Research on customer lifetime value based on machine learning algorithms and customer relationship management analysis model," Heliyon, vol. 9, no. 2, e13384, 2023, doi: 10.1016/j.heliyon.2023.e13384.

12. K. Gadgil, S. S. Gill, and A. M. Abdelmoniem, "A meta-learning based stacked regression approach for customer lifetime value prediction," Journal of Economy and Technology, vol. 1, pp. 197-207, 2023, doi: 10.1016/j.ject.2023.09.001.
13. Y. Yan and N. Resnick, "A high-performance turnkey system for customer lifetime value prediction in retail brands," Quantitative Marketing and Economics, vol. 22, no. 2, pp. 169-192, 2024, doi: 10.1007/s11129-023-09272-x.
14. E. B. Firmansyah, M. R. Machado, and J. L. R. Moreira, "How can artificial intelligence (AI) be used to manage customer lifetime value (CLV)? A systematic literature review," International Journal of Information Management Data Insights, vol. 4, no. 2, 100279, 2024, doi: 10.1016/j.jjimei.2024.100279.
15. A. A. Tudoran, C. H. Thomsen, and S. Thomasen, "Understanding consumer behavior during and after a pandemic: Implications for customer lifetime value prediction models," Journal of Business Research, vol. 174, 114527, 2024, doi: 10.1016/j.jbusres.2024.114527.
16. S. S. Poudel, S. Pokharel, and M. Timilsina, "Explaining customer churn prediction in telecom industry using tabular machine learning models," Machine Learning with Applications, vol. 17, 100567, 2024, doi: 10.1016/j.mlwa.2024.100567.
17. S.-V. Oprea and A. Bâra, "Customer-centric decision-making with XAI and counterfactual explanations for churn mitigation," Journal of Theoretical and Applied Electronic Commerce Research, vol. 20, no. 2, art. 129, 2025, doi: 10.3390/jtaer20020129.
18. O. Doğan, A. Hızıroğlu, A. Pişirgen, and O. F. Seymen, "Business analytics in customer lifetime value: An overview analysis," WIREs Data Mining and Knowledge Discovery, vol. 15, no. 1, e1571, 2025, doi: 10.1002/widm.1571.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

