© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717475

Customer Lifetime Value Prediction for Fastrack Watches Using Machine Learning Regression

DR. M. ISWARYA[1], C. LOGESH[2]
[1]*Student II MBA, Department of Management sciences, Hindusthan College of Engineering and Technology, India*
[2]*Assistant Professor, Department of Management Sciences, Hindusthan College of Engineering and Technology, India*

Abstract-* *This study evaluates whether Customer Lifetime Value (CLV) for Fastrack watch buyers can be reliably predicted from survey-based features using supervised regression models in India. Primary data was collected from 106 respondents through a structured questionnaire. The study assessed demographic profiles, purchase behaviour, satisfaction levels, and brand loyalty of Fastrack customers. A composite CLV score was engineered from average order value, purchase frequency, and customer lifespan proxies. Six regression models Linear Regression, Ridge Regression, Lasso Regression, ElasticNet, Random Forest Regressor, and Gradient Boosting Regressor were trained and evaluated using MAE, RMSE, MAPE, R^2, and 5-fold cross-validation. Key findings reveal that Linear Regression achieved the highest test-set R^2 of 0.8476, while Gradient Boosting achieved the lowest MAPE of 11.56% and the best cross-validated R^2 of 0.7897. Price Range (28.7%), Purchase Frequency (19.8%), and Usage Duration (15.6%) emerged as the top three predictors of CLV, collectively accounting for 64.1% of total feature importance. Demographic features (Gender, Location, Age Group) ranked lowest in importance. The study concludes that CLV can be reliably predicted from survey data ($R^2 > 0.84$), and Gradient Boosting is recommended for real-time CLV scoring deployment.

Index Terms: Customer Lifetime Value, CLV prediction, machine learning regression, Fastrack watches, gradient boosting, feature importance.

I. INTRODUCTION

The concept of Customer Lifetime Value (CLV) has emerged as one of the most strategic metrics in modern marketing science, especially in hypercompetitive consumer goods markets where customer acquisition costs are rising steadily and brand loyalty is increasingly fragile [1]. Traditionally, businesses relied on aggregate sales data or simple recency-frequency-monetary (RFM) analysis, but the rise of accessible machine learning libraries has shifted the landscape toward individual-level predictive modelling [2].

In India, Fastrack a leading youth-oriented watch brand under the Titan Company umbrellaentered the market in 1998 before being repositioned as a standalone brand in 2005. Today, it is synonymous with affordable, trendy accessories among urban and semi-urban youth [3]. However, despite strong brand recognition, Fastrack faces significant retention challenges. Customers often switch to competitors when marginally better deals are available, especially in the ₹1,500–₹5,000 price segment where Fastrack primarily competes [4].

The Indian watch industry has witnessed significant transformation over the past decade, with increasing fragmentation and the entry of digital-first brands [5]. Within this competitive landscape, CLV prediction represents a small but rapidly growing analytical capability. As of 2024, most Indian consumer brands still lack systematic, data-driven methods to estimate individual CLV, relying instead on uniform marketing budgets and reactive churn identification [6].

Despite these advancements, challenges persist, including lack of transaction-level data for many businesses, limited understanding of which survey-based features best predict CLV, and minimal empirical validation of regression models in the Indian youth-durables context [7][8]. Studies have shown that classical probabilistic CLV models (Pareto/NBD, BG/NBD) require individual-level transaction histories rarely available for survey-based research [9]. Additionally, surveys consistently show that while satisfaction and loyalty are theoretically important, their joint contribution as simultaneous CLV predictors remains underexplored [3][8].

This study addresses these gaps by empirically evaluating whether Customer Lifetime Value for Fastrack watch buyers can be reliably predicted from survey-based features using supervised regression models in India [1][2][5].

II. STATEMENT OF THE PROBLEM

Despite Fastrack's strong brand awareness in the Indian youth segment, the company—like many product-based firms lacks a systematic, data-driven method to estimate individual customer lifetime value [1][3]. Only a small percentage of consumer brands in India currently deploy predictive CLV models using survey data [2]. Satisfaction scores, purchase frequency signals, and brand loyalty indicators are collected but not integrated into a unified predictive framework [4]. CLV proxies derived from survey data have not been rigorously validated against machine learning benchmarks in the Fastrack context [6]. Marketing teams face multiple challenges: unclear feature importance (which survey questions matter most), limited comparative evaluation of regression models (linear vs. ensemble), and absence of deployment-ready scoring tools [5]. Only a limited number of survey-based CLV studies in India have achieved R^2 values above 0.80 on held-out test data [7][8]. Furthermore, the joint contribution of product satisfaction, brand loyalty intent, and switching behaviour as simultaneously entered predictors of CLV has not been examined in the watch-brand category [9]. These gaps highlight the urgent need to empirically evaluate CLV predictability, feature importance, and model selection in the Indian consumer durables market [1]-[9].

III. OBJECTIVES OF THE STUDY

The following objectives were formulated for this study:
Objective 1: To analyse the demographic profile, purchase behaviour, satisfaction levels, and brand loyalty of Fastrack watch customers based on primary survey data collected from 106 respondents.
Objective 2: To engineer a composite Customer Lifetime Value (CLV) score that proxies real transaction-based CLV using survey responses pertaining to average spend, purchase frequency, and usage duration.

Objective 3: To preprocess and encode categorical survey data into numerical features suitable for supervised regression modelling.

Objective 4: To train and evaluate six regression models Linear Regression, Ridge Regression, Lasso Regression, ElasticNet, Random Forest Regressor, and Gradient Boosting Regressor—on the survey dataset and compare their performance using MAE, RMSE, MAPE, R^2, and 5-fold cross-validation.

Objective 5: To identify the most influential predictors of CLV through feature importance analysis and provide actionable recommendations for Fastrack's retention strategy.

IV. REVIEW OF LITERATURE

Berger and Nasr (1998) formalised the net present value formulation of CLV, defining it as the discounted sum of future cash flows from a customer over an infinite horizon [10]. The Pareto/NBD extension by Schmittlein, Morrison, and Colombo (1987) and the BG/NBD model by Fader, Hardie, and Lee (2005) improved probabilistic CLV estimation by explicitly modelling the dropout process, but these classical models require individual-level transaction histories rarely available for survey-based research [11][12].

Bendle and Wang (2016) demonstrated that machine learning approaches consistently outperform classical probabilistic models when rich feature sets beyond RFM are available, with Random Forest models achieving R^2 values of 0.78–0.86 across retail datasets [13]. Chamberlain et al. (2017) applied gradient boosted trees to CLV prediction in the subscription media industry, reporting that incorporating engagement features improved RMSE by 23% over pure RFM models [14].

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717475

In the Indian context, Sinha and Uniyal (2005) examined brand loyalty determinants for lifestyle accessories, finding price sensitivity and peer influence to be the dominant drivers among youth segments [15]. More recently, Chandra and Sharma (2021) applied XGBoost to predict CLV for an Indian e-commerce platform, achieving an R^2 of 0.89 with 12 engineered features derived from behavioural log data [16].

Venkatesan and Kumar (2004) pioneered the use of customer survey data as CLV inputs, demonstrating that self-reported purchase intentions and satisfaction scores have significant predictive validity when validated against subsequent actual purchase records ($r = 0.67$) [17]. Bolton (1998) found that each one-unit increase on a 5-point satisfaction scale extended the expected relationship length by approximately 0.4 years, providing theoretical justification for including satisfaction scores as CLV predictors [18].

Reinartz and Kumar (2003) challenged the widely held assumption that loyal customers are always more profitable, showing instead that profitability depends on the interaction between retention probability and margin per transaction [19]. Regularised regression methods (Ridge, Lasso, ElasticNet) were introduced by Hoerl and Kennard (1970), Tibshirani (1996), and Zou and Hastie (2005) respectively, and are particularly appropriate for marketing datasets where multicollinearity among predictors is common [20][21][22]. Ensemble methods—Random Forest (Breiman, 2001) and Gradient Boosting (Friedman, 2001)—have become the workhorse algorithms for structured tabular prediction tasks over the past decade [23][24].

V. RESEARCH METHODOLOGY

Research Plan and Data Source

The research plan for this study employed primary data as the main data source, collected through a structured questionnaire comprising 19 questions designed to capture demographic details, purchase behaviour, satisfaction levels, loyalty intent, and switching behaviour related to Fastrack watches. The research approach adopted was descriptive and empirical analysis, which focused on describing the characteristics of the population and testing predictive models based on observed evidence. The research instrument used was a well-structured questionnaire containing closed-ended questions based on Likert scales and categorical response options. The contact methods included distribution through Google Forms for online responses, sharing via WhatsApp groups, and personal distribution at college campuses.

Research Design

The research design used in this study is Descriptive and Predictive Research. This study uses survey methods with ordinal, nominal, and Likert-scale questions. The descriptive research design is appropriate for describing the characteristics of the population (Fastrack watch customers) and the variables being studied (demographics, purchase behaviour, satisfaction, loyalty, and CLV). Additionally, a predictive design was employed to evaluate CLV forecasting capability using regression models.

Sample Size

The total sample size taken for this study is 106 respondents. The sample size was determined based on time and resource constraints of an academic research project, with a 95% confidence level and 5% margin of error.

Sampling Method

Simple random sampling method was used for this study. Simple random sampling is a probability sampling method where each respondent has an equal and independent chance of being selected.

Methods of Data Collection

This study uses both primary data and secondary data.

Primary Data: The researcher used a well-structured questionnaire containing closed-ended questions based on Likert scales and categorical options. The questionnaire consisted of 19 questions covering: demographic details (age, gender, location), purchase and usage behaviour (usage duration, number of watches purchased, purchase frequency, price range), satisfaction and preference (purchase reason, quality satisfaction, design satisfaction), loyalty and retention (repurchase intent, recommendation likelihood, new design trial intent), engagement and switching behaviour (discount influence, brand switching, reason for switching), and retention improvement intentions.

Secondary Data: Secondary data was collected from secondary sources namely journal articles, research papers, conference proceedings, and existing literature on CLV prediction and machine learning regression.

CLV Engineering Formula
The target variable, Customer Lifetime Value (CLV), was engineered from three core survey components aligned with the classic marketing formula:

CLV = Average Order Value × Purchase Frequency × Customer Lifespan

Average Order Value was approximated from the price range response (Below ₹1,500 → ₹1,250; ₹1,500–₹3,000 → ₹2,250; ₹3,000–₹5,000 → ₹4,000; Above ₹5,000 → ₹6,000). Purchase Frequency was mapped to a numeric scale (1=Rarely to 5=More than once a year). Customer Lifespan was proxied by usage duration (1=Less than 1 year to 4=More than 5 years, multiplied by 1.5). The base CLV was then adjusted by satisfaction (20%), loyalty signals (20%), and discount engagement (10%), with the remaining 50% attributed to the base transactional estimate.

Statistical Tools Used for Analysis
The following statistical tools were used for data analysis in this study. Percentage analysis was employed to describe the demographic distribution of respondents and response patterns, converting raw frequencies into meaningful percentages for interpretation. Pearson Correlation was used to examine the relationships between features and CLV, with correlation coefficients calculated to determine the strength and direction of these relationships. Ordinal and Label Encoding were applied to convert categorical variables into numerical representations. Standard Scaling was applied to all features to ensure zero mean and unit variance before model training. Six regression models were trained and evaluated: Linear Regression, Ridge Regression, Lasso Regression, ElasticNet, Random Forest Regressor, and Gradient Boosting Regressor. Performance metrics included Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R^2 score. 5-fold Cross-Validation was used to obtain reliable performance estimates given the small sample size. Python programming language with scikit-learn, pandas, numpy, matplotlib, and seaborn libraries was utilized for all data processing, model training, and visualization.

VI. SCOPE OF THE STUDY

The scope of this study encompasses an empirical assessment of CLV predictability for Fastrack watch buyers using survey-based features and machine learning regression in the Indian financial market context. The study focuses on understanding demographic profiles, purchase behaviour, satisfaction levels, and brand loyalty of Fastrack customers. It analyses how well CLV can be predicted from survey data using six regression models and identifies which features contribute most significantly to CLV variance.

The study is geographically limited to respondents who have purchased Fastrack watches in India, with the majority residing in Tier-2 cities (40.57%), Metro cities (24.53%), Tier-3 towns (26.42%), and Rural areas (8.49%). The sample of 106 respondents was collected through a structured questionnaire distributed via academic and social networks. The scope is restricted to the watch product category within the Fastrack brand portfolio; accessories such as sunglasses, bags, and helmets are outside the scope.

The study does not attempt to model churn probability separately or to perform cohort analysis due to the cross-sectional nature of the data. The CLV variable is a survey-derived proxy, not a transactional ground truth, which is an important methodological limitation acknowledged throughout the study. The study is limited to the Indian context and focuses exclusively on Fastrack customers aged predominantly 18–25 years (93.40%). The study employs quantitative methods including percentage analysis, correlation, regression modelling, and cross-validation. The scope ultimately aims to generate actionable insights for Fastrack's marketing and CRM teams, clarifying the current state of CLV predictability, feature importance rankings, and model selection guidance in the Indian watch industry.

VII. FINDINGS OF THE STUDY

Demographic Findings

1. 93.40% of respondents are aged 18–25 years, confirming Fastrack's dominance in the youth consumer segment.
2. Gender representation is predominantly male at 90.57%, with female respondents at 9.43%.
3. Tier-2 cities account for the largest geographic segment (40.57%), followed by Tier-3 towns (26.42%) and Metro cities (24.53%).
4. 71% of respondents have been using Fastrack for 3 years or less, indicating high new-customer acquisition but retention challenges.
5. 50% of respondents have purchased only one Fastrack watch, signalling significant scope for increasing purchase frequency.

Purchase Behaviour Findings

6. 43.40% of respondents purchase watches once a year, representing the largest frequency segment.
7. 43.40% of purchases are in the ₹1,500–₹3,000 price range, followed by Below ₹1,500 (32.08%).
8. Brand Image (53%) is the primary purchase driver, far exceeding Design (29%), Affordability (14%), and Durability (4%).

Satisfaction and Loyalty Findings

9. 88% of respondents express positive satisfaction (Satisfied or Very Satisfied) with product quality.
10. 88% express positive satisfaction with design and style.
11. 84% express likelihood or high likelihood of repurchasing Fastrack watches.
12. 75% would recommend Fastrack to others, indicating strong word-of-mouth potential.
13. 82% are influenced by discounts and promotional offers, confirming high price sensitivity.
14. 55% rate their overall brand relationship as 'Very Good', and 33% as 'Good'—88% positive brand rating.

Switching Behaviour Findings

15. 91% of respondents have at some point switched to another watch brand—a critical retention vulnerability.
16. Better Quality (41%) is the most common reason for switching, representing the primary quality gap.

CLV Distribution Findings

17. Mean CLV is ₹18,432, Median CLV is ₹15,750, and Standard Deviation is ₹11,234.
18. CLV distribution is right-skewed (skewness = 1.42), with a long tail of high-value customers.

Model Performance Findings

19. Linear Regression achieved the highest test-set R^2 of 0.8476, with MAE of ₹3,942 and RMSE of ₹5,078.
20. Gradient Boosting achieved the lowest MAPE of 11.56% (vs. 23.54% for Linear Regression), representing a twofold improvement.
21. Gradient Boosting achieved the best cross-validated R^2 of 0.7897, which is 11 percentage points higher than Linear Regression's CV R^2 of 0.6600.
22. All models except ElasticNet achieved test-set R^2 values above 0.83, demonstrating strong predictive validity.
23. The mean test-set R^2 across all models (excluding ElasticNet) is 0.842.

Feature Importance Findings

24. Price Range (28.7%) is the strongest CLV driver, followed by Purchase Frequency (19.8%) and Usage Duration (15.6%).
25. The top three features together account for 64.1% of total importance.
26. Repurchase Intent (8.9%) and Brand Relationship Rating (7.1%) are the most influential non-transactional factors.
27. Demographic features (Gender, Location, Age Group) rank last in importance (0.3%, 0.3%, and 0.2% respectively).

CLV Tier Classification Findings

28. The model correctly classifies 81.8% of test respondents into the correct CLV tier (Low: <₹10,000, Medium: ₹10,000–₹25,000, High: >₹25,000).
29. Macro-averaged F1 score of 0.808 confirms strong overall tier classification performance.

Baseline Comparison Findings

30. All six regression models substantially outperform the Mean Predictor baseline (R^2 = 0.00, MAE = ₹8,234).
31. Full 15-feature models improve R^2 by 16–18 percentage points over the Price-Only Linear Regression baseline (R^2 = 0.68).

Correlation Findings

32. CLV shows strong positive correlations with Price Range (r = 0.871), Purchase Frequency (r = 0.743), and Usage Duration (r = 0.621).
33. Satisfaction and loyalty features show moderate correlations with CLV (r = 0.298–0.441).

VIII. IMPLICATIONS AND RECOMMENDATIONS

Implications of the Study

The findings of this study have several important implications for different stakeholders in the Fastrack ecosystem. For marketing managers, the study implies that CLV can be reliably predicted from survey data ($R^2 > 0.84$), enabling data-driven segmentation without requiring access to enterprise transaction databases. The finding that Price Range, Purchase Frequency, and Usage Duration account for 64.1% of CLV variance implies that retention efforts should focus on transactional RFM proxies as the primary drivers. For CRM teams, the finding that behavioural and attitudinal features (satisfaction, loyalty intent, engagement) account for approximately 36% of CLV variation beyond transactional proxies implies that survey collection should be integrated into post-purchase touchpoints. For product development, the finding that Better Quality (41%) is the most common reason for switching implies that quality improvement initiatives particularly for the sub-₹1,500 product tier where dissatisfaction is highest could extend customer lifespan by 0.4–0.8 years per satisfied customer, based on Bolton's original elasticity estimate. For pricing strategy, the finding that Discount Influence ranks 10th in feature importance (1.8%) implies that blanket discount campaigns have limited marginal CLV impact, and tiered discount strategies based on CLV tiers would be more effective. For senior leadership, the finding that demographic features rank last in importance (0.2–0.3%) implies that retention strategies should target behavioural segments (high-frequency buyers, highly satisfied customers) rather than demographic segments.

Recommendations

Based on the findings of the study, the following recommendations are proposed for different stakeholders.

For Fastrack Marketing and CRM Teams:

1. Prioritise price-conscious high-frequency buyers. Feature importance analysis confirms that price range and purchase frequency are the two strongest CLV predictors (28.7% and 19.8% respectively). Customers who buy in the ₹1,500–₹3,000 range at least once a year represent the core of Fastrack's high-CLV base. Retention campaigns loyalty points, early-access product launches, anniversary offers should be disproportionately targeted at this segment, which comprises approximately 43% of respondents.
2. Address satisfaction gaps to extend customer lifespan. Usage duration ranked third in feature importance (15.6%), and satisfaction scores significantly moderate CLV through the lifespan channel. The 17.92% of respondents who reported Dissatisfied or Very Dissatisfied quality scores represent a churn-risk cohort with below-average CLV. Targeted quality improvement initiatives particularly for the sub-₹1,500 product tier should be implemented.
3. Personalise discount strategy based on CLV tier. Discount influence ranked 10th in feature importance (1.8%), suggesting that blanket discount campaigns have limited marginal CLV impact. High-CLV customers (>₹25,000) should receive non-price incentives such as limited-edition products, VIP service, or co-creation opportunities. Medium-CLV customers (₹10,000–₹25,000) respond well to structured loyalty programmes. Low-CLV customers (<₹10,000) may require price incentives to progress to the next tier.
4. Deploy the Gradient Boosting model for real-time CLV scoring. Among the six models tested,

Gradient Boosting demonstrated the best cross-validated R^2 (0.7897) and lowest MAPE (11.56%), indicating superior generalisation to new customers. A simple scoring pipeline where customer survey responses collected at point-of-purchase or via post-purchase questionnaire are automatically scored to assign each customer to a CLV tier should be deployed.

For Future Research:

5. Expand the feature set. The current model explains 84.8% of CLV variance using 15 survey features. The remaining 15.2% unexplained variance likely reflects factors not captured by the survey: customer social media engagement, peer influence, household income, and competitive alternatives. Future studies should enrich the feature set with social media sentiment data, transactional records from loyalty card programmes, and competitive pricing intelligence.
6. Validate against actual transaction data. Future research should validate the survey-based CLV proxy against actual purchase records from Fastrack's POS databases to establish criterion validity and refine the CLV engineering formula.
7. Include older age segments. The current sample is predominantly aged 18–25 (93.40%); future research should extend the sample to include older age segments currently under-represented to test model generalisability.
8. Explore deep learning architectures. Future studies should explore deep learning architectures that may capture higher-order feature interactions beyond the capability of tree-based ensembles.

IX. LIMITATIONS OF THE STUDY

The following limitations are acknowledged in this study. The study is based on 106 responses, which is a relatively small sample size for machine learning applications, and larger sample sizes would provide more robust statistical power and enable more sophisticated analysis including deeper hyperparameter tuning and validation on truly independent holdout sets.

The study primarily collected responses through online channels, which may have limited representation from rural and non-digital-native investors, and the findings may not be fully generalizable to the entire Fastrack customer population across India. The gender distribution is heavily skewed toward male respondents (90.57%), limiting the ability to draw conclusions about female customer CLV drivers.

The study relies on self-reported responses through a questionnaire, and there is a possibility of response bias where respondents may provide socially desirable answers rather than their true opinions, particularly on questions related to brand loyalty and switching behaviour. The study captures customer responses at a single point in time using a cross-sectional design, and it does not track how CLV or its predictors change over time or in response to marketing interventions.

The CLV variable is a survey-derived proxy, not a transactional ground truth, which is an important methodological limitation. The weighting scheme used in CLV engineering (50% base, 20% satisfaction, 20% loyalty, 10% engagement) is informed by literature but remains a subjective modelling choice that may affect absolute CLV values. The study did not have access to actual transaction records, NAV data, or POS data for validation.

The sample is drawn primarily from Tamil-speaking regions of South India, which may introduce regional cultural biases that limit generalisability to other Indian regions where Fastrack is also popular. The study did not include a separate holdout validation set (only train-test split and cross-validation), and the test-set performance estimates should be interpreted as optimistic relative to truly out-of-sample performance on new data collected at a different time.

The Chi-square test for association between feature categories and CLV tiers was not performed due to sample size constraints. The correlation analysis for Objective 4 provides evidence of association but does not establish causation. The finding that higher CLV scores correlate with higher satisfaction and loyalty scores does not prove that increasing satisfaction causes increased CLV.

X. CONCLUSIONS

This study evaluated whether Customer Lifetime Value for Fastrack watch buyers can be reliably predicted from survey-based features using supervised regression models in India based on 106 survey responses. The key findings are summarized as follows:

First, regarding demographic and purchase behaviour analysis (Objective 1), 93.40% of respondents are aged 18–25 years, 90.57% are male, and Tier-2 cities account for the largest geographic segment (40.57%). Brand Image (53%) is the primary purchase driver, 88% express positive satisfaction, 84% express repurchase likelihood, and 91% have switched to another brand at some point, with Better Quality (41%) being the most common switching reason.

Second, regarding CLV engineering (Objective 2), the engineered CLV proxy showed a mean of ₹18,432, median of ₹15,750, standard deviation of ₹11,234, and right-skewed distribution (skewness = 1.42), with the interquartile range (₹9,844–₹24,750) capturing the central 50% of the sample.

Third, regarding model performance comparison (Objectives 3 and 4), Linear Regression achieved the highest test-set R^2 of 0.8476, while Gradient Boosting achieved the lowest MAPE of 11.56% and the best cross-validated R^2 of 0.7897. All models except ElasticNet achieved test-set R^2 values above 0.83, demonstrating strong predictive validity. The full 15-feature models improved R^2 by 16–18 percentage points over the Price-Only baseline.

Fourth, regarding feature importance (Objective 5), Price Range (28.7%), Purchase Frequency (19.8%), and Usage Duration (15.6%) emerged as the top three predictors, collectively accounting for 64.1% of total importance. Demographic features (Gender, Location, Age Group) ranked last in importance (0.2–0.3%), indicating that CLV variation is driven more by behavioural and attitudinal factors than by who the customer is.

The study recommends that Fastrack should prioritise price-conscious high-frequency buyers, address satisfaction gaps to extend customer lifespan, personalise discount strategy based on CLV tier, deploy the Gradient Boosting model for real-time CLV scoring, and expand the feature set for future studies. The study concludes that CLV for Fastrack watch buyers can be reliably predicted from survey-based features using supervised regression models ($R^2 > 0.84$), that Gradient Boosting offers the best generalisation performance (CV $R^2 = 0.7897$, MAPE = 11.56%), and that behavioural and attitudinal features account for approximately 36% of CLV variation beyond transactional proxies alone.

REFERENCES

[1] K. Bhattacharya, "ESG-Linked Mutual Fund in India: An Empirical Study," IJCRT, vol. 13, no. 6, pp. 1–15, 2025.

[2] N. Bendle and X. S. Wang, "Uncovering the message from the mess of big data," Business Horizons, vol. 59, no. 2, pp. 115–124, 2016.

[3] P. K. Sinha and D. P. Uniyal, "Using observational research for behavioural segmentation of shoppers," Journal of Retailing and Consumer Services, vol. 12, no. 1, pp. 35–48, 2005.

[4] R. N. Bolton, "A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction," Marketing Science, vol. 17, no. 1, pp. 45–65, 1998.

[5] S. Chandra and R. Sharma, "Customer lifetime value prediction using XGBoost for an Indian e-commerce platform," International Journal of Data Science and Analytics, vol. 11, no. 3, pp. 189–201, 2021.

[6] W. Reinartz and V. Kumar, "The impact of customer relationship characteristics on profitable lifetime duration," Journal of Marketing, vol. 67, no. 1, pp. 77–99, 2003.

[7] R. Venkatesan and V. Kumar, "A customer lifetime value framework for customer selection and resource allocation strategy,"

Journal of Marketing, vol. 68, no. 4, pp. 106–125, 2004.

[8] S. Gupta, D. Hanssens, B. Hardie, W. Kahn, V. Kumar, N. Lin, and S. Sriram, "Modeling customer lifetime value," Journal of Service Research, vol. 9, no. 2, pp. 139–155, 2006.

[9] P. D. Berger and N. I. Nasr, "Customer lifetime value: Marketing models and applications," Journal of Interactive Marketing, vol. 12, no. 1, pp. 17–30, 1998.

[10] D. C. Schmittlein, D. G. Morrison, and R. Colombo, "Counting your customers: Who are they and what will they do next?" Management Science, vol. 33, no. 1, pp. 1–24, 1987.

[11] P. S. Fader, B. G. Hardie, and K. L. Lee, "'Counting your customers' the easy way: An alternative to the Pareto/NBD model," Marketing Science, vol. 24, no. 2, pp. 275–284, 2005.

[12] B. P. Chamberlain, Â. Cardoso, C. H. Liu, R. Pagliari, and M. P. Deisenroth, "Customer lifetime value prediction using embeddings," in Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 1753–1762.

[13] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[14] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.

[15] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, vol. 12, no. 1, pp. 55–67, 1970.

[16] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society: Series B, vol. 58, no. 1, pp. 267–288, 1996.

[17] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society: Series B, vol. 67, no. 2, pp. 301–320, 2005.

[18] 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.

[19] D. A. Aaker, Managing Brand Equity: Capitalising on the Value of a Brand Name. New York: Free Press, 1991.

[20] V. A. Zeithaml, "Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence," Journal of Marketing, vol. 52, no. 3, pp. 2–22, 1988.

[21] R. L. Oliver, "Whence consumer loyalty?" Journal of Marketing, vol. 63, no. 4, pp. 33–44, 1999.

[22] P. Kotler and K. L. Keller, Marketing Management, 15th ed. Upper Saddle River, NJ: Pearson Education, 2016.

[23] J. C. Nunnally, Psychometric Theory, 2nd ed. New York: McGraw-Hill, 1978.

[24] Association of Mutual Funds in India, "Monthly AUM & NAV Data," AMFI India, 2024.