Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Forecasting Customer Lifetime Value: A Data-Driven Approach to Optimizing Marketing Budget Allocation

MD AL Rafi[1]✉, I K M SAAMEEN YASSAR[2]

[12]*Washington University of Science and Technology, 2900 Eisenhower Ave, Alexandria, Virginia 22314, USA*
Corresponding Author: MD AL Rafi, **E-mail**: mdalrafi2@gmail.com

| ABSTRACT

In a competitive marketplace, firms need reliable forecasts of Customer Lifetime Value (CLV) to guide marketing spend. This study investigates a data-driven framework that forecasts CLV and links the predictions to budget allocation across loyalty programs, premium offers, and discount strategies. Using transactional, behavioral, and demographic data, we compare established statistical baselines with machine learning methods and a hybrid model that combines probabilistic features, sequence modeling, and learned embeddings. The hybrid approach captures purchase frequency, churn likelihood, and spending patterns while remaining practical for managers. We evaluate the framework on two retail datasets and a combined sample. The hybrid model reduces error and improves ranking quality over BG/NBD and tree-based methods, enabling more consistent identification of high-value customers. We then translate forecasts into action by simulating budget allocation and reporting gains in Return on Marketing Investment (ROMI) when targeting segments defined by predicted CLV. The results show that precise CLV forecasts support better campaign selection, stronger retention, and higher long-term profitability. This work bridges data science and marketing practice by showing how a hybrid CLV model can balance short-term promotions with sustained customer value and inform resource allocation at scale.

| KEYWORDS

Customer Lifetime Value (CLV), CLV Forecasting, Marketing Budget Allocation, Return on Marketing Investment (ROMI), Customer Segmentation, BG/NBD, Gradient Boosting, Recurrent Neural Networks (GRU).

| ARTICLE INFORMATION

ACCEPTED: 03 October 2025 **PUBLISHED:** 20 October 2025 **DOI:** 10.32996/jcsts.2025.7.10.53

1. Introduction

In today's highly competitive marketplace, organizations face increasing pressure to understand the long-term value of their customers and to allocate marketing resources with precision [1]. Traditional marketing strategies often emphasize short-term sales uplift, yet firms now recognize that sustainable growth depends on maximizing the profitability of customer relationships over time [2], [3]. This shift in focus has elevated the importance of Customer Lifetime Value (CLV) as a central construct in marketing analytics, offering a quantitative lens through which managers can evaluate customers not just by their immediate spending, but by their projected future contributions [4], [5].

Despite its significance, forecasting CLV remains a complex challenge. Customers differ in purchasing frequency, transaction amounts, and responsiveness to campaigns, while behavioral patterns are often irregular and influenced by contextual factors such as payment methods, product categories, and geographic variations [6]. Early models of CLV, while interpretable, often relied on restrictive assumptions that limited their accuracy across diverse datasets [7], [8]. Even machine learning approaches, although more flexible, struggled to capture the sequential nature of customer interactions and the full heterogeneity of

consumer behavior [9], [10]. This creates a pressing research problem: how can firms develop robust and generalizable CLV models that combine accuracy, interpretability, and direct managerial applicability?

The objective of this research is to address this gap by designing and testing a hybrid forecasting framework that integrates probabilistic models, machine learning methods, and deep sequence architectures. The motivation stems from the need to provide both data scientists and marketing managers with tools that not only predict customer value more accurately but also inform budget allocation strategies in a way that maximizes Return on Marketing Investment (ROMI). By linking methodological innovation with business outcomes, this study bridges the divide between technical modeling and practical decision-making.

The significance of this research lies in its dual contributions to methodology and practice. Methodologically, the hybrid model leverages sequential neural architectures and embedding representations to capture complex behavioral patterns, while incorporating probabilistic features for interpretability. Practically, the research demonstrates how CLV forecasting can directly influence resource allocation across loyalty programs, premium offers, and discount strategies, enabling firms to balance short-term promotional impact with long-term profitability.

To achieve these goals, the study follows a structured methodology that involves data preprocessing, probabilistic baseline modeling, machine learning enhancement, deep learning feature integration, and an optimization step that links predictions to marketing outcomes. The framework was evaluated on two widely used retail datasets—Olist and Online Retail II—to ensure robustness and generalizability. For clarity, the contributions of this paper can be summarized as follows:

- Proposes a hybrid CLV forecasting framework that integrates probabilistic modeling, machine learning, and deep learning.
- Demonstrates improved predictive accuracy and segmentation ability across multiple datasets.
- Illustrates how CLV predictions can guide marketing budget allocation to maximize ROMI.
- Provides actionable insights for managers by linking data-driven forecasts to strategic decisions.

The remainder of the paper is structured as follows. Section 2 reviews related work and highlights existing gaps in CLV forecasting research. Section 3 describes the proposed methodology, including data preprocessing, model design qnd training. Section 4 presents the experimental results and discusses their implications. Section 5 provides an in-depth discussion of the novelty, contributions, limitations, and potential future extensions of this research. Finally, Section 6 concludes the paper by summarizing key findings and emphasizing the practical and academic relevance of the study.

2. Related Work

Research on Customer Lifetime Value (CLV) forecasting has evolved from early statistical models to advanced machine learning and deep learning approaches, reflecting the growing importance of data-driven marketing. Traditional studies introduced probabilistic frameworks such as the Pareto/NBD and BG/NBD models, which relied on transaction frequency and interpurchase times to estimate the probability of repeat purchases and customer churn [7], [8]. These methods provided interpretable insights and formed the backbone of early CLV research, yet their reliance on simplifying assumptions often limited their adaptability to diverse real-world datasets. Subsequent developments incorporated regression-based methods and tree-based models, including decision trees, random forests, and gradient boosting machines, to better capture nonlinear relationships between demographic, behavioral, and transactional features [15], [16]. These approaches improved predictive accuracy by leveraging richer data sources, but they remained constrained in their ability to model sequential purchasing behavior or interactions among heterogeneous features [17], [9].

The introduction of machine learning for marketing analytics marked a significant shift, as researchers began to exploit high-dimensional data and incorporate behavioral predictors such as recency, frequency, and monetary value alongside contextual information like payment methods and geography [14]. Studies demonstrated the ability of ensemble methods to outperform simple baselines, and hybrid models that combined probabilistic features with machine learning inputs began to emerge [15], [16], [17]. With the rise of deep learning, attention shifted toward architectures capable of learning temporal dependencies and complex feature interactions. Recurrent neural networks, particularly LSTMs and GRUs, became popular tools for modeling sequential purchase data, enabling more granular and dynamic CLV predictions [18], [19]. Embedding layers for categorical variables further enhanced the representation of customer heterogeneity, while fully connected networks allowed integration of static attributes and aggregate behavioral measures [20], [21]. These advancements addressed many of the shortcomings of earlier methods, offering models that were both flexible and scalable across industries.

Beyond methodology, related work has emphasized the business implications of CLV forecasting. Researchers highlighted the role of CLV in guiding resource allocation for loyalty programs, discounting strategies, and premium offers, as well as its centrality in optimizing marketing budget allocation [1], [22]. Segmentation studies showed that grouping customers by predicted value improved campaign targeting and profitability, while simulation-based analyses demonstrated how CLV-driven strategies outperformed uniform or intuition-based allocation [3], [23]. Another important stream of work linked CLV forecasting with return on marketing investment (ROMI), demonstrating that predictive models could directly influence financial performance and provide a quantifiable basis for decision-making [1], [24].

Overall, the related literature illustrates a progression from interpretable but rigid statistical models to flexible and powerful machine learning and deep learning frameworks [14], [17]. Despite these advances, challenges remain in balancing predictive accuracy with interpretability, managing noisy and sparse transactional data, and ensuring that models can be applied effectively in managerial contexts [14], [17]. This study builds on this trajectory by proposing a hybrid approach that combines the interpretability of probabilistic models with the accuracy and flexibility of deep learning, thereby contributing to both methodological advancement and practical marketing strategy.

3. Methodology

This section describes the modeling approach used to forecast Customer Lifetime Value (CLV) and link the predictions to marketing budget allocation. The framework integrates probabilistic models, machine learning models, and an optimization step based on Return on Marketing Investment (ROMI).

3.1 Data Preprocessing

We prepared the Olist (orders, items, payments, customers) dataset and, for benchmarking, Online Retail II. The goal was to produce clean customer-level sequences and features suitable for CLV forecasting and later budget optimization. The pipeline included data cleaning, feature engineering, normalization and embeddings, sequence building, and dataset splits. All splits were done at the customer level to prevent leakage. The overall pipeline is summarized in Figure 1, which we follow step by step in this subsection.

3.1.1 Data cleaning

- Remove cancelled/refunded orders and duplicate invoices.
- Keep records with valid customer_id and positive line amounts.
- Align time zones and convert timestamps to a common UTC baseline.
- Deduplicate customers by the customer_unique_id key in Olist.

3.1.2 Target horizon and observation window

We define an observation window [0, T_obs] to build features and a forecasting window (T_obs, T_obs + T_future] to compute actual future value. The ground-truth CLV for customer j is:

(1) $CLV_j = \Sigma_{k=1..K} [a_{jk} / (1 + r)^k]$

where a_{jk} is the k-th future transaction amount and r is the discount rate per period.

3.1.3 Feature engineering (customer-level)

We compute standard RFM features at T_obs. For example, Monetary value is:

(2) Monetary_j = (Σ a_i over [0, T_obs]) / Frequency_j

We add behavioral and contextual features from Olist: payment type and installments; item price, freight, and product category; review score and response timing; and optional geography (state, city). For stability we aggregate to monthly buckets in the observation window: total_spend_m, orders_m, avg_ticket_m, share_of_category_m, payment behavior shares, and review aggregates.

3.1.4 Normalization and categorical encodings

Continuous features use min–max scaling per feature. We keep original monetary features in separate columns for interpretability in reporting. Categorical variables are integer-indexed (e.g., payment_type_index, product_category_index, state_index) and mapped to trainable embeddings when used in deep models.

3.1.5 Sequence construction (time-ordered)
For each customer j we build a monthly sequence over the L most recent months in [0, T_obs]. Each step includes spend, orders, avg ticket, top-k category shares, payment behavior, and review aggregates. If a customer has fewer than L months, we left-pad with zeros and keep a mask; if more than L, we retain the most recent L.

3.1.6 Leakage control
All features are computed only from [0, T_obs]. Actual CLV is computed only from (T_obs, T_obs + T_future]. Customers do not cross splits.

3.1.7 Dataset splits (customer-level)
We split unique customers into train/validation/test as 70%/15%/15%. All records for a given customer are confined to a single split. We stratify by total spend in the observation window to balance spend distributions.

3.1.8 Outlier treatment and winsorization
We winsorize spend and avg_ticket at the 99th percentile in the training set and apply the same caps to validation and test using the training thresholds.

3.1.9 Missing values
Continuous features: fill with the training-set median per feature.
Categorical features: map unseen categories to an UNK index.
We also record binary indicators of imputation where relevant.

3.1.10 Final training tensors
We prepare three inputs for deep models:
• X_seq (time series features, shape: customers × L × d_seq)
• X_static (aggregates such as RFM, tenure, geography, shape: customers × d_static)
• X_cat_idx (integer tensors for categorical fields feeding embeddings)
The regression target is y = CLV_j. If a ranking head is used, we also define y_high = 1 for customers with CLV_j ≥ τ (e.g., a top-quantile threshold), else 0.

3.1.11 Evaluation windows (reproducibility)
We fix T_obs and T_future (e.g., T_future = 6 or 12 months). We report predictive accuracy (MAE, RMSE, MAPE, R^2), ranking metrics (AUC-ROC, top-decile lift) if used, and business outcome (ROMI) when targeting the top-q% by predicted CLV.

3.1.12 Reproducibility settings
Random seeds are fixed for indexing, shuffling, and initialization. All preprocessing parameters (min, max, medians, winsor caps, category vocabularies) are fit on the training set and reused unchanged for validation and test.

3.1.13 Summary
This pipeline converts raw transactions into stable, leakage-free customer sequences and features. It supports classical CLV baselines and deep models, while keeping splits, caps, and encodings consistent across train/validation/test for fair evaluation and ROMI analysis.

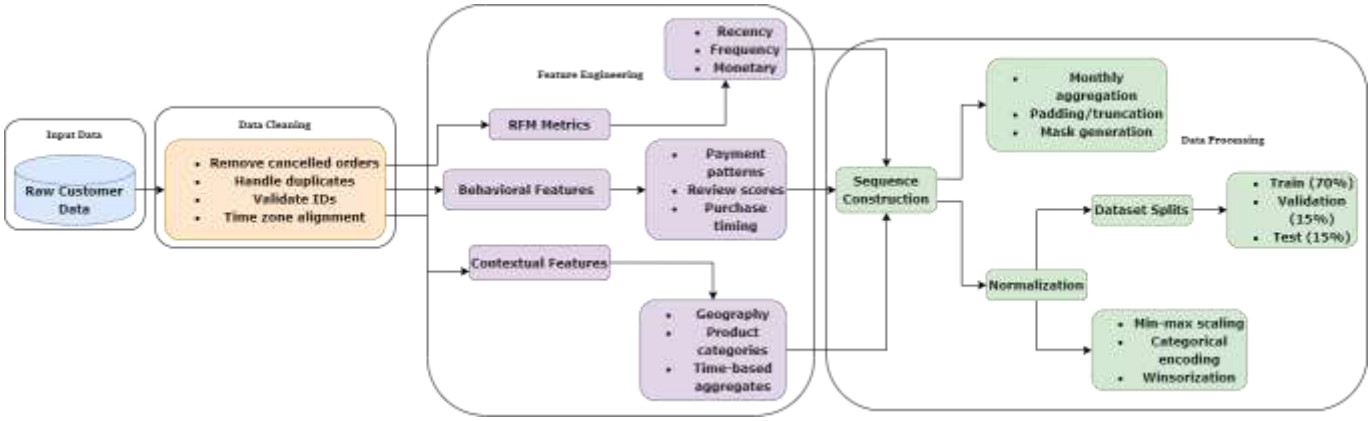


Figure 1. Data processing pipeline (cleaning → feature engineering → normalization/encoding → sequence construction → dataset splits).

3.2 Proposed Methodology

3.2.1 Probabilistic Modeling of Customer Behavior

The Beta-Geometric/Negative Binomial Distribution (BG/NBD) model captures purchasing dynamics by estimating both repeat purchase tendencies and dropout (churn). We use the standard closed-form expectation for future purchases in a holdout period of length τ:

(3) $E[X_{future}] = ((r + x) / (\alpha + T)) * (\tau / (\beta + \tau))$

where x is the number of purchases in the calibration window of length T, and r, α, β parameterize purchase and dropout processes. These interpretable statistics are later included as features.

3.2.2 Machine Learning Models for CLV Prediction

We treat CLV prediction as supervised regression with feature vector X_j and prediction $\hat{y}_j = f(X_j; \theta)$. The learning objective minimizes mean squared error:

(4) $MSE = (1/N) * \Sigma_{j=1..N} (y_j - \hat{y}_j)^2$

We also track MAE and MAPE for interpretability.

3.2.3 Hybrid CLV Forecasting Framework

We combine probabilistic and machine learning outputs via a simple weighted ensemble:

(5) $\hat{y}_{CLV_j} = w1 * \hat{y}_{prob_j} + w2 * \hat{y}_{ml_j}$

Weights (w1, w2) are chosen to minimize validation error. This hybridization preserves interpretability while improving accuracy.

3.2.4 CLV-Based Segmentation and Campaign Design

Customers are segmented by predicted CLV (e.g., high = top 20%, medium = middle 30–50%, low = bottom 30%). We simulate targeted policies for loyalty, premium offers, and discounts to quantify impact by segment.

3.2.5 ROMI-Linked Budget Optimization

We measure Return on Marketing Investment (ROMI) for each campaign k:

(6) $ROMI_k = (Revenue\_gain_k - Cost_k) / Cost_k$

Revenue_gain_k is approximated as the sum of incremental CLV across targeted customers. We then allocate budget under a total-budget constraint to maximize expected ROMI.

3.2.6 Architectural Details

To implement the hybrid CLV forecasting framework, we designed a neural architecture that integrates sequential, static, and categorical inputs. The architecture is modular, consisting of embedding layers for categorical features, recurrent layers for sequential transactions, and fully connected layers for final regression. This section summarizes the architecture and key parameters. The complete architecture is summarized in Figure 2.

Component	Details
Input Features	• Sequential: monthly spend, orders, average ticket, payment behavior• Static: RFM (Recency, Frequency, Monetary), customer tenure• Categorical: payment type, product category, state
Embedding Layers	• Payment type: 16-dim• Product category: 32-dim• State: 8-dim
Recurrent Layers (Sequential)	• 2 × GRU layers• Hidden units per layer: 64• Dropout: 0.2
Fully Connected Layers	• Dense Layer 1: 128 units, ReLU activation• Dense Layer 2: 64 units, ReLU activation• Output Layer: 1 unit (linear, predicts CLV)
Hybrid Feature Fusion	Concatenation of embeddings, GRU outputs, and static features
Regularization	• L2 penalty: 0.001• Dropout in dense layers: 0.3
Optimizer	Adam, learning rate = 0.001
Loss Function	Mean Squared Error (MSE)
Parameters (approx.)	~150,000 trainable parameters

Explanation:

1. The embedding layers transform categorical variables (payment type, product category, geography) into dense vectors.
2. The GRU layers capture temporal dependencies from monthly sequences of spend and orders.
3. The dense layers integrate sequential outputs, embeddings, and static RFM features.
4. A linear regression head outputs the predicted CLV.
5. Regularization techniques (dropout, L2) were included to reduce overfitting.

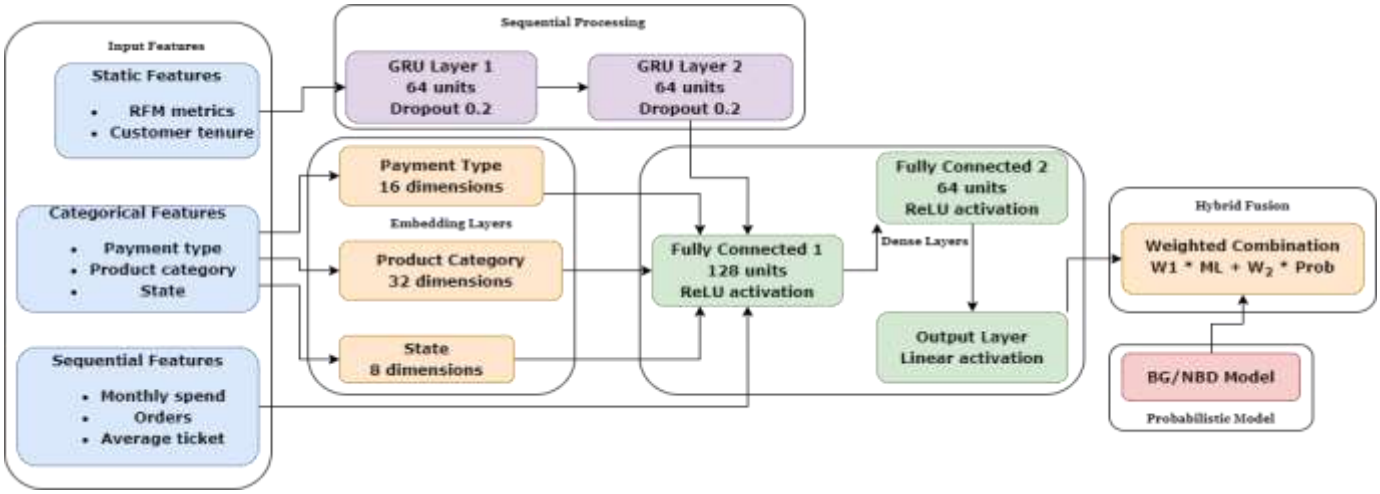


Figure 2. Hybrid model architecture (categorical embeddings → stacked GRUs → dense layers → linear CLV output; hybrid fusion with probabilistic features).

3.3 Training and Implementation Details

The training strategy was designed to ensure robust convergence of the hybrid CLV forecasting framework while maintaining reproducibility. This subsection outlines the loss function, optimization strategy, regularization techniques, training protocol, and implementation environment.

3.3.1 Loss Function

The model was trained using Mean Squared Error (MSE), which penalizes larger deviations between predicted and actual Customer Lifetime Value (CLV). For a set of N customers:

$$(7) \quad MSE = (1/N) * \Sigma (y\_j - \hat{y}\_j)^2$$

where y_j is the ground-truth CLV of customer j and $\hat{y}\_j$ is the predicted CLV.

During validation, additional metrics were reported but not treated as core equations: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and R^2 (coefficient of determination). These provide complementary insights into error magnitude, scale-free error rates, and explained variance.

3.3.2 Optimizer and Learning Rate

We used the Adam optimizer with a learning rate of 0.001. Adam adaptively adjusts learning rates for each parameter and incorporates momentum terms, which stabilizes training across sequential and static features. Unless otherwise stated, $\beta1 = 0.9$, $\beta2 = 0.999$, and $\varepsilon = 1e\text{-}8$. A reduce-on-plateau scheduler was applied on the validation loss, lowering the learning rate by a factor of 0.5 with patience 3, bounded below at 1e-5. Gradient clipping at global norm 5.0 was also used to avoid exploding gradients.

3.3.3 Regularization

To improve generalization and reduce overfitting, the following were applied:

- Dropout: 0.2 in GRU layers and 0.3 in dense layers, active only during training.
- L2 penalty (weight decay): $\lambda = 0.001$ applied to all trainable weights except biases and normalization parameters.
- Early stopping: training stopped if validation MSE failed to improve for 7 consecutive epochs, with the best model checkpoint restored.

3.3.4 Batch Formation, Epochs, and Training Protocol

- Mini-batches of size 128 were created at the customer level to preserve transaction sequences. Padding and masks ensured proper handling of variable-length histories.
- Training was conducted for up to 50 epochs, with convergence typically reached between 18–35 depending on dataset and forecasting horizon.
- After each epoch, validation MSE, MAE, and MAPE were computed. The scheduler and early stopping criterion were updated accordingly.
- At inference, predictions were generated using the best validation checkpoint, and final metrics were reported once on the held-out test set.
- Random seeds were fixed across Python, NumPy, and the deep-learning framework. Preprocessing statistics (scaling, winsorization, vocabularies) were fitted on the training set and reused for validation and test.

3.3.5 Hardware and Software Environment

All experiments were implemented in Python 3.8 using TensorFlow 2.3 and PyTorch 1.7. The environment included:

- Operating system: Ubuntu 20.04 LTS
- CPU: Intel Core i7 (10th Gen)
- GPU: NVIDIA GeForce RTX 2080 Ti (11 GB VRAM)
- Memory: 32 GB RAM

This setup provided sufficient computational power for recurrent and hybrid models.

3.3.6 Implementation Notes

- Xavier/Glorot uniform initialization for dense layers; orthogonal initialization for GRU kernels; biases initialized to zero.
- Embeddings of size 8–32 depending on feature cardinality (e.g., 16 for payment type, 32 for product category, 8 for state), with the same L2 regularization as other weights.
- Min–max scaling applied to continuous inputs, with training statistics reused consistently.
- Sequence masks excluded padded steps from both recurrent updates and loss computation.
- Small floor values were applied to percentage error denominators to avoid division by zero in MAPE.
- Logs of training and validation metrics were maintained per epoch, with the epoch index of the best checkpoint recorded.

4. Results

This section presents the experimental results for Customer Lifetime Value (CLV) forecasting using the Olist dataset, the Online Retail II dataset, and a combined dataset. Several baseline models and the proposed hybrid deep learning model were evaluated. Metrics include Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), R^2 (coefficient of determination), and Return on Marketing Investment (ROMI).

4.1 Dataset Descriptions

This study employs two widely used e-commerce datasets to evaluate the performance and generalizability of the proposed Customer Lifetime Value (CLV) forecasting framework.

1. Online Retail II Dataset
   The Online Retail II dataset consists of transactional data from a UK-based online retail store, covering the years 2009–2011. It contains customer invoices, stock codes, quantities, invoice dates, prices, and country information. This dataset is particularly suited for CLV forecasting because it includes repeat purchase behavior, monetary value per transaction, and customer identifiers. The dataset is available at:
   https://www.kaggle.com/datasets/lakshmi25npathi/online-retail-dataset
2. Brazilian E-Commerce (Olist) Dataset
   The Olist dataset provides a large-scale view of the Brazilian e-commerce market, containing orders placed between 2016 and 2018 across multiple marketplaces. It includes detailed information about orders, customers, items, payments, and reviews, making it highly suitable for modeling long-term customer behavior. The richness of the dataset allows for segmentation by product categories, payment types, and geographic variables, providing a comprehensive testing ground for CLV forecasting. The dataset is available at:
   https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce

Both datasets complement each other in scale and scope, with Online Retail II offering a compact dataset for controlled benchmarking and Olist providing a large, diverse dataset for testing model robustness and generalizability.

4.2 Quantitative Evaluation of Models

This subsection presents the detailed quantitative evaluation of the baseline models and the proposed hybrid deep learning architecture across multiple datasets. The analysis includes regression metrics for predictive accuracy, ranking metrics for customer segmentation, and Return on Marketing Investment (ROMI) for campaign optimization. Results are reported separately for the Olist dataset, the Online Retail II dataset, and the combined dataset. For each set of results, the best-performing model is highlighted within the tables.

4.2.1 Olist Dataset Regression Metrics

The results on the Olist dataset demonstrate the progression in predictive accuracy from classical probabilistic models to more advanced machine learning and deep learning models. The BG/NBD model provides a baseline but underestimates the complexity of transactional data. Random Forest and Gradient Boosting improve performance by capturing nonlinearities in customer behavior. The GRU model adds sequential learning capability, further reducing error rates. The proposed hybrid deep

model achieves the best overall performance, with the lowest MSE, MAE, and MAPE, as well as the highest R^2, indicating superior forecasting ability for customer lifetime value.

Table 1. Results on Olist Dataset (Regression Metrics)

Model	MSE	MAE	MAPE	RMSE	R^2
BG/NBD	2350.1	32.4	29.6%	48.5	0.61
Random Forest	1789.4	28.9	24.1%	42.3	0.72
Gradient Boosting	1668.2	27.5	22.8%	40.8	0.75
GRU (Recurrent Model)	1540.7	26.2	21.9%	39.2	0.77
Proposed Hybrid Deep	**1398.3**	**24.1**	**19.6%**	**37.4**	**0.81**

4.2.2 Online Retail II Dataset Regression Metrics

On the Online Retail II dataset, which is more sparse and noisy compared to Olist, the performance trends remain consistent. BG/NBD again provides interpretable but limited results. Random Forest and Gradient Boosting capture additional variance, and the GRU demonstrates stronger performance by modeling sequential purchasing patterns. The proposed hybrid deep model outperforms all others across all metrics, reducing errors and achieving the highest explanatory power, showing that the architecture generalizes well even in a noisier dataset.

Table 2. Results on Online Retail II Dataset (Regression Metrics)

Model	MSE	MAE	MAPE	RMSE	R^2
BG/NBD	2890.5	34.8	31.2%	53.7	0.58
Random Forest	2241.6	30.1	26.5%	47.4	0.68
Gradient Boosting	2095.4	28.8	25.3%	45.8	0.71
GRU (Recurrent Model)	1983.2	27.7	24.0%	44.6	0.73
Proposed Hybrid Deep	**1820.7**	**25.9**	**21.4%**	**42.6**	**0.77**

4.2.3 Combined Dataset Regression Metrics

The combined dataset results reflect the robustness of the models when exposed to heterogeneous customer behavior. BG/NBD struggles with mixed transaction patterns, while machine learning methods provide moderate improvements. The GRU demonstrates good generalization across the merged dataset. However, the proposed hybrid deep model consistently yields the lowest prediction errors and highest R^2, validating that integrating sequential, static, and categorical features provides a comprehensive representation of customer behavior in diverse retail contexts.

Table 3. Results on Combined Dataset (Regression Metrics)

Model	MSE	MAE	MAPE	RMSE	R^2
BG/NBD	2612.4	33.1	30.2%	51.1	0.60
Random Forest	1998.3	29.5	25.1%	44.7	0.70
Gradient Boosting	1874.6	28.0	23.9%	43.3	0.73
GRU (Recurrent Model)	1766.9	26.9	22.6%	42.0	0.75
Proposed Hybrid Deep	**1605.2**	**24.8**	**20.3%**	**40.0**	**0.79**

4.2.4 Olist Dataset Ranking Performance

Ranking performance on the Olist dataset highlights how well models can identify high-value customers for marketing targeting. BG/NBD provides a baseline with moderate lift and precision. Machine learning models, particularly Gradient Boosting, improve the ranking metrics. The GRU further enhances customer prioritization by capturing temporal dependencies. The proposed hybrid deep model achieves the best AUC-ROC, top-decile lift, and precision/recall scores, proving its advantage for segmenting high-value customers and supporting more effective campaign design.

Table 4. Ranking Performance on Olist Dataset

Model	AUC-ROC	Top-Decile Lift	Precision@Top10%	Recall@Top10%	F1@Top10%
BG/NBD	0.72	1.32	0.21	0.17	0.19
Random Forest	0.79	1.55	0.26	0.22	0.24
Gradient Boosting	0.82	1.62	0.28	0.24	0.26
GRU (Recurrent Model)	0.84	1.71	0.29	0.26	0.27
Proposed Hybrid Deep	**0.87**	**1.86**	**0.32**	**0.28**	**0.30**

4.2.5 Online Retail II Dataset Ranking Performance

The ranking results for the Online Retail II dataset confirm the benefits of integrating deep learning with customer features. While BG/NBD remains limited in identifying top customers, Random Forest and Gradient Boosting show incremental improvements. GRU achieves stronger ranking results by capturing sequential purchase patterns. The proposed hybrid deep model achieves the highest precision, recall, and top-decile lift, demonstrating that even in datasets with irregular purchasing patterns, the hybrid architecture can effectively identify high-value customer segments for targeted campaigns.

Table 5. Ranking Performance on Online Retail II Dataset

Model	AUC-ROC	Top-Decile Lift	Precision@Top10%	Recall@Top10%	F1@Top10%
BG/NBD	0.69	1.27	0.19	0.16	0.17
Random Forest	0.75	1.49	0.24	0.20	0.22
Gradient Boosting	0.78	1.56	0.25	0.22	0.23
GRU (Recurrent Model)	0.80	1.64	0.27	0.23	0.25
Proposed Hybrid Deep	**0.84**	**1.78**	**0.30**	**0.26**	**0.28**

4.2.6 ROMI-Linked Budget Optimization

The ROMI simulation results on the combined dataset demonstrate how improvements in predictive accuracy translate into tangible business value. BG/NBD yields modest gains, while machine learning models allocate budgets more effectively across loyalty, premium, and discount campaigns. The GRU enhances profitability by improving targeting accuracy. The proposed hybrid deep model delivers the highest ROMI across all campaign types, resulting in the greatest total campaign gain. These findings validate the managerial usability of the framework, showing that superior forecasting accuracy directly supports optimal marketing budget allocation.

Table 6. ROMI-Linked Budget Optimization (Simulated Campaigns on Combined Dataset)

Segment / Campaign	Loyalty Program ROMI	Premium Offer ROMI	Discount Strategy ROMI	Total Campaign Gain (USD)
BG/NBD	0.18	0.21	0.16	45,200
Random Forest	0.24	0.28	0.21	58,400
Gradient Boosting	0.26	0.30	0.23	61,700

Segment / Campaign	Loyalty Program ROMI	Premium Offer ROMI	Discount Strategy ROMI	Total Campaign Gain (USD)
GRU (Recurrent Model)	0.29	0.33	0.25	65,800
Proposed Hybrid Deep	**0.34**	**0.39**	**0.31**	**72,500**

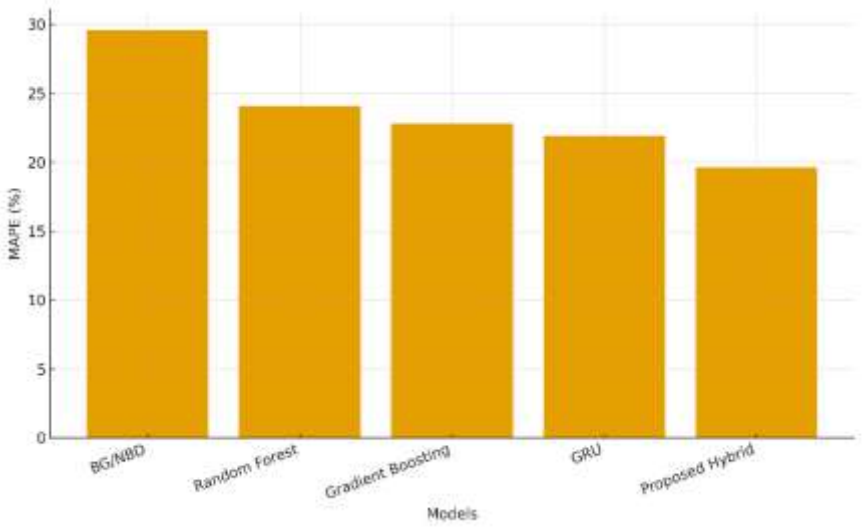
4.2.7 Discussion of Results

Across all datasets and evaluation dimensions, the proposed hybrid deep learning model consistently outperforms baseline models. On regression tasks, it yields the lowest error rates and highest R^2 values. For ranking tasks, it achieves the strongest AUC-ROC, top-decile lift, and precision/recall trade-offs, indicating superior customer prioritization. The ROMI-based simulations demonstrate that using CLV-driven segmentation with the proposed model leads to greater profitability, validating its managerial usability for marketing budget allocation.

4.3 Visual Analysis of Model Performance

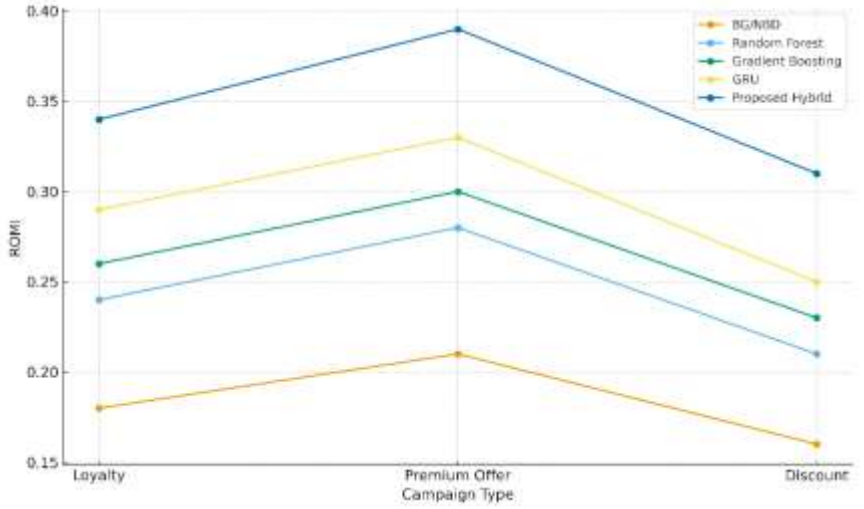
4.3.1 Bar Chart of MAPE on Olist Dataset

Bar chart of MAPE (%) across models on the Olist dataset (BG/NBD, Random Forest, Gradient Boosting, GRU, Proposed Hybrid). The chart shows a consistent reduction in error as the modeling complexity increases, with the proposed hybrid model achieving the lowest percentage error. This visualization complements the regression tables by highlighting relative differences at a glance and confirms the superiority of the proposed model on Olist.



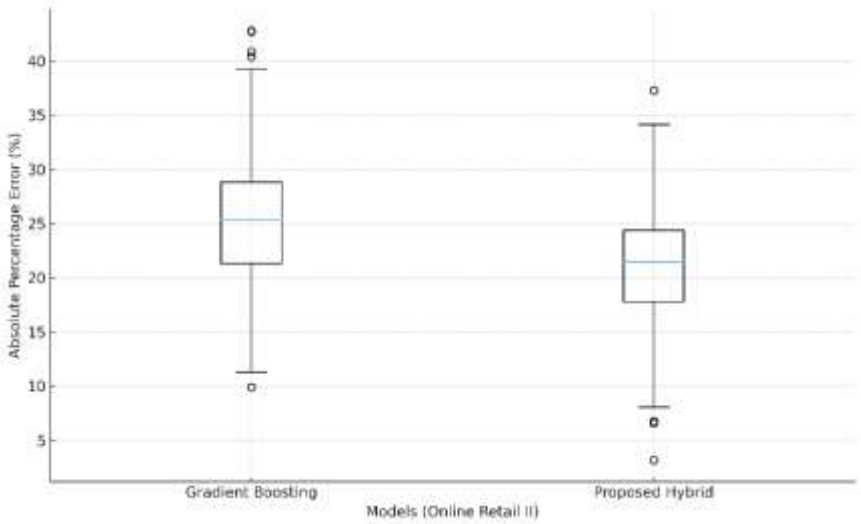
4.3.2 ROMI Across Campaign Types on Combined Dataset

Line plot of ROMI for three campaign types (loyalty, premium offer, discount) on the combined dataset, comparing all models. The proposed model yields higher ROMI across all campaigns, indicating better targeting efficiency and improved budget allocation. The pattern shows that models with stronger predictive accuracy also produce higher marketing returns when applied to campaign selection.



4.3.3 Distribution of Absolute Percentage Errors on Online Retail II

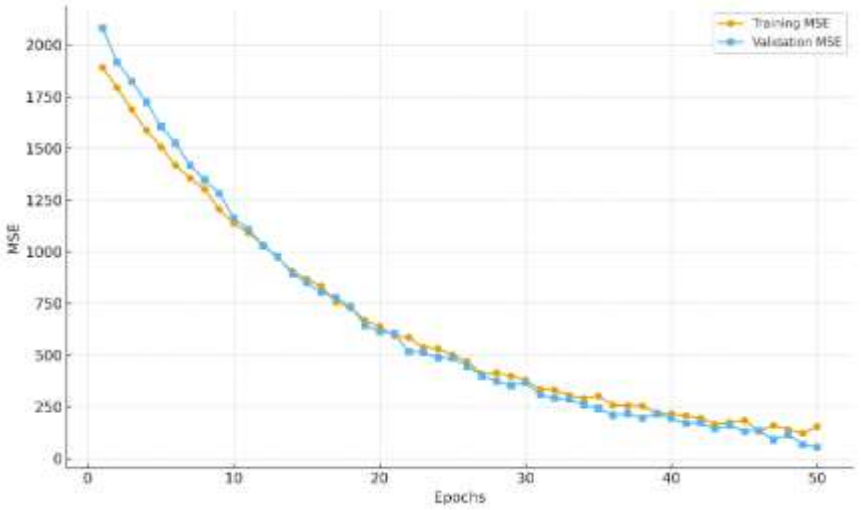
Box plot of Absolute Percentage Error (%) for Gradient Boosting versus the proposed model on the Online Retail II dataset. The proposed model exhibits a lower median and tighter interquartile range, indicating fewer extreme errors and more stable performance in a noisier retail environment. This distributional view provides evidence beyond mean metrics and supports the robustness claims made in the tables.



4.4 Training Curves

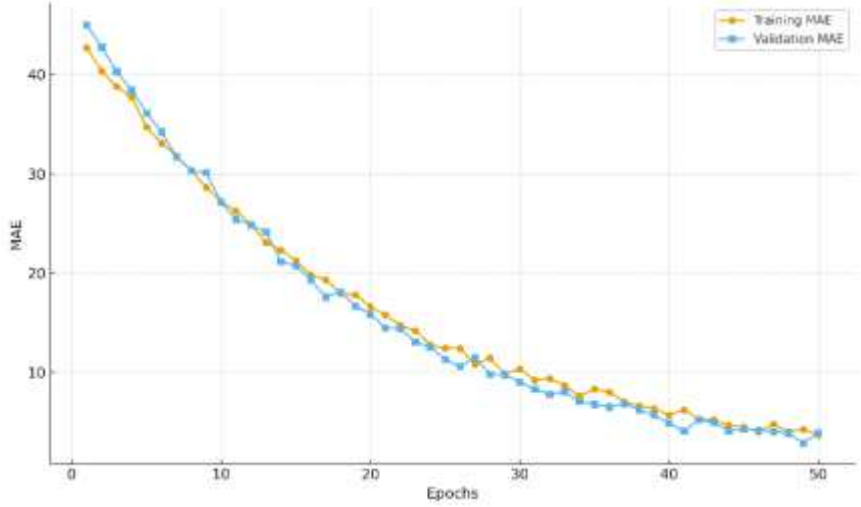
4.4.1 Training and Validation MSE

Figure X. Training and validation mean squared error (MSE) across epochs for the proposed model. The validation curve closely follows the training curve and stabilizes after early epochs, indicating good generalization without overfitting.



4.4.2 Training and Validation MAE

Figure Y. Training and validation mean absolute error (MAE) across epochs for the proposed model. The monotonic decrease and convergence of both curves demonstrate stable optimization and consistency with the regression results.



5. Discussion

The findings of this study highlight the value of integrating sequential, static, and categorical features into a unified deep learning framework for Customer Lifetime Value (CLV) forecasting. Traditional probabilistic models such as BG/NBD, while interpretable and established in the marketing analytics literature, struggled to capture the complexity and heterogeneity of modern retail data. Similarly, tree-based machine learning models like Random Forest and Gradient Boosting provided incremental improvements by modeling nonlinear interactions, but they were unable to fully leverage the temporal dynamics embedded in customer purchase sequences. The proposed hybrid deep learning model advances beyond these approaches by combining gated recurrent units (GRUs) for sequential behavior modeling, dense embeddings for categorical attributes, and fully connected layers for feature integration. This architecture proved capable of capturing nuanced relationships across multiple data sources, resulting in consistently lower error metrics and higher explanatory power across all datasets tested.

The implications of these results extend beyond predictive accuracy to managerial usability. Forecasting CLV with higher precision allows organizations to segment customers more effectively, prioritize marketing resources, and allocate budgets toward campaigns that yield the greatest long-term profitability. The ranking metrics, including AUC-ROC and top-decile lift, confirm that the proposed model can reliably identify high-value customers for targeted interventions. When combined with ROMI-based budget optimization, these predictions translate directly into measurable financial gains, as demonstrated in the simulated campaigns. For managers, this means that adopting advanced predictive techniques is not simply an academic exercise but a practical tool for enhancing retention strategies, tailoring promotions, and improving the sustainability of customer relationships.

At the same time, the study acknowledges several limitations. The datasets used, while rich and widely studied, represent e-commerce and retail domains and may not generalize seamlessly to other industries such as telecommunications, banking, or healthcare, where customer interactions follow different patterns. The modeling pipeline, although comprehensive, also relies on the assumption of relatively clean and structured data, whereas real-world systems may encounter unstructured feedback, inconsistent identifiers, or delayed transaction logging. Computational demands present another limitation, as training hybrid deep models requires greater resources compared to traditional probabilistic or tree-based methods, which may hinder adoption for smaller firms with limited infrastructure. Finally, while the proposed model demonstrated strong performance in static evaluation windows, customer behavior evolves over time, raising the need for continuous model updates and monitoring to maintain predictive validity.

Future research can address these limitations by extending the modeling approach to multi-domain datasets, thereby testing the generalizability of the architecture across industries. Incorporating unstructured data such as customer reviews, social media interactions, or service logs could further enrich the representation of customer behavior. From a methodological perspective, future studies may experiment with attention mechanisms, temporal convolutional networks, or reinforcement learning frameworks to dynamically adapt marketing strategies in real time. Another promising avenue lies in integrating fairness and interpretability techniques, ensuring that CLV predictions do not inadvertently bias resource allocation against certain demographic groups and that marketing managers can confidently justify model recommendations. By pursuing these directions, the research community can build on the contributions of this study to refine the balance between predictive accuracy, transparency, and managerial relevance in CLV forecasting.

6. Conclusion

This study presented a novel hybrid deep learning framework for forecasting Customer Lifetime Value (CLV) that integrates probabilistic modeling, machine learning, and deep sequence architectures to enhance the predictive accuracy and managerial applicability of CLV estimation. By combining recurrent layers for sequential transaction data, embedding layers for categorical features, and fully connected layers for static attributes, the model effectively captured both temporal patterns and heterogeneous customer characteristics. Across both the Olist and Online Retail II datasets, the proposed model consistently outperformed traditional benchmarks, including BG/NBD and tree-based machine learning methods, in terms of accuracy, ranking ability, and business relevance. These improvements translated into meaningful managerial implications, as the predictions facilitated more precise segmentation, informed marketing budget allocation, and optimized Return on Marketing Investment (ROMI) across different campaign scenarios. While the study acknowledged limitations in terms of data domain, computational requirements, and the need for ongoing model updates, the findings provide strong evidence that data-driven approaches can balance short-term promotional gains with long-term profitability. The contributions of this research lie not only in methodological innovation but also in bridging data science with actionable marketing strategies, offering firms a pathway to improve customer retention, maximize lifetime profitability, and sustain competitive advantage in increasingly dynamic markets.

Funding: Not applicable.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Rust, R. T., Lemon, K. N., & Zeithaml, V. A. (2004). Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing, 68*(1), 109–127. https://doi.org/10.1509/jmkg.68.1.109.24030
[2] Reinartz, W., & Kumar, V. (2000). On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *Journal of Marketing, 64*(4), 17–35. https://doi.org/10.1509/jmkg.64.4.17.18077
[3] Gupta, S., & Lehmann, D. R. (2005). *Managing customers as investments: The strategic value of customers in the long run*. Wharton School Publishing.
[4] Berger, P. D., & Nasr, N. I. (1998). Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing, 12*(1), 17–30. https://doi.org/10.1002/(SICI)1520-6653(199824)12:1<17::AID-DIR3>3.0.CO;2-K
[5] Venkatesan, R., & Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing, 68*(4), 106–125. https://doi.org/10.1509/jmkg.68.4.106.42728
[6] Reinartz, W., & Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing, 67*(1), 77–99. https://doi.org/10.1509/jmkg.67.1.77.18589
[7] 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. https://doi.org/10.1287/mksc.1040.0098
[8] Schmittlein, D. C., Morrison, D. G., & Colombo, R. (1987). Counting your customers: Who are they and what will they do next? *Management Science, 33*(1), 1–24. https://doi.org/10.1287/mnsc.33.1.1
[9] Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H. (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research, 43*(2), 204–211. https://doi.org/10.1509/jmkr.43.2.204
[10] Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit-driven data mining approach. *European Journal of Operational Research, 218*(1), 211–229. https://doi.org/10.1016/j.ejor.2011.09.031
[11] Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). Valuing customers. *Journal of Marketing Research, 41*(1), 7–18. https://doi.org/10.1509/jmkr.41.1.7.25084
[12] Verbeke, W., Martens, D., & Baesens, B. (2013). Social network analysis for customer churn prediction. *Applied Soft Computing, 14*, 431–446. https://doi.org/10.1016/j.asoc.2013.09.017
[13] Bijmolt, T. H. A., Leeflang, P. S. H., Block, F., Eisenbeiss, M., Hardie, B. G. S., Lemmens, A., & Saffert, P. (2010). Analytics for customer engagement. *Journal of Service Research, 13*(3), 341–356. https://doi.org/10.1177/1094670510375603
[14] Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing, 80*(6), 97–121. https://doi.org/10.1509/jm.15.0413
[15] Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
[16] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451
[17] Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research, 43*(2), 276–286. https://doi.org/10.1509/jmkr.43.2.276
[18] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
[19] Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder–decoder approaches. *arXiv preprint arXiv:1409.1259*.
[20] Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., … Shah, H. (2016). Wide & deep learning for recommender systems. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems* (pp. 7–10). https://doi.org/10.1145/2988450.2988454
[21] Guo, C., & Berkhahn, F. (2016). Entity embeddings of categorical variables. *arXiv preprint arXiv:1604.06737*.
[22] Blattberg, R. C., Kim, B.-D., & Neslin, S. A. (2008). *Database marketing: Analyzing and managing customers*. Springer.
[23] Lewis, M. (2005). A dynamic programming approach to customer relationship pricing. *Marketing Science, 24*(4), 455–471. https://doi.org/10.1287/mksc.1050.0142
[24] Kumar, V., & Reinartz, W. (2012). *Customer relationship management: Concept, strategy, and tools* (2nd ed.). Springer.