Proceedings of the 3rd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2025.28947

Privacy-Preserving Rolling CLV Forecast for Smarter Budget Allocation

Wenli Huang

Coventry University, Coventry, UK
huang77sandra@gmail.com

Abstract. In the current data-driven digital economy environment, customer lifetime value (CLV) prediction is becoming the basis for enterprises to formulate precise marketing strategies and optimize budget allocation. However, the traditional centralized modeling approach faces challenges of information leakage and regulatory restrictions, while the unpredictability of the market requires the prediction model to have real-time adaptability. Therefore, static modeling cannot meet the requirements of real-time decision-making. Thus, this study proposes an idea for a rolling CLV prediction framework, which combines differential privacy with federated learning, enabling secure prediction of customer value and optimizing dynamic budget allocation in the scenario of multi-store cooperation. By using public data from offline experience stores of Suning, Gome and JD from 2019 to 2023, this study establishes a federated learning framework under the protection of the differential privacy mechanism, captures the dynamic changes in customer behavior using a rolling time window, and designs an intelligent budget allocation strategy based on the CLV prediction results. Experimental results confirm that the proposed framework achieves a 91.2% prediction effect under strict confidentiality restrictions ($\varepsilon = 1.0$), with a 53.4% improvement in prediction accuracy compared to traditional strategies, and an increase in the return on investment from 1.9 to 4.2, and a 9.7% increase in customer retention rate. This study provides an innovative privacy-protecting solution in a multi-party cooperation scenario, with significant theoretical value and practical significance, and helps to regulate the interpretability development of retail enterprises' marketing.

Keywords: Differential Privacy, Federated Learning, Rolling Forecast, Customer Lifetime Value (CLV), Budget Allocation

1. Introduction

In the context of the data-driven digital economy that is oriented towards big data, CLV prediction has become an important foundation for enterprises to build specific marketing strategies and optimize budget allocation. Through quantitative modeling of customer transaction behaviors, enterprises can not only identify high-value customers but also achieve more efficient market investment under limited resources. However, most current methods fundamentally rely on centralized data aggregation and training. Although the prediction accuracy has significantly improved, they have led to information leakage and compliance issues [1]. With the implementation

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.28947

of information protection regulations, the conflict between the availability of information and compliance has become increasingly prominent, making traditional centralized modeling face sustainability challenges. At the same time, the uncertainty of the market environment requires prediction models to have dynamic adaptability, and fixed models cannot meet the requirements of real-time decision-making [2]. Therefore, this study proposes a rolling prediction framework that combines differential privacy with federated learning. This framework can achieve federated learning for the protected customer lifetime value prediction in the scenario of multi-store cooperation and realize dynamic optimization of budget allocation through rolling updates. This research opens up a new technical path that integrates information protection and intelligent marketing, and provides practical guidance.

2. Literature review

2.1. Advances in customer lifetime value forecasting

Customer lifetime value (CLV) prediction, as a component of the comprehensive tool for evaluating long-term customer contributions, has undergone an evolution process from early models based on statistical regression and RFM indicators to multi-stage methods based on machine learning and deep learning. Traditional models rely on parameter assumptions, which can handle the relationship between customer behavior and their value. However, they have obvious limitations when dealing with high-dimensional and complex behavioral data [3]. In recent decades, various techniques including random forest, gradient boosting tree, and recurrent neural networks were used for CLV forecasting. Such approaches possess strong abilities in processing nonlinearity and time-series problems [4]. Nonetheless, despite the promising predicting capacity of these models from the experimental perspective, strong presumptions regarding the availability of big, comprehensive and centralized data pose many restrictions [5]. In addition to that, the variation in customer distribution in various outlets is another factor that limits the universality of the model and makes the stability of predictions questionable in decision-making-based scenarios.

2.2. Privacy-preserving computational methods

With the continuous strengthening of the data protection law, differential privacy and federated learning have emerged as the two major approaches to tackle problems with the modeling of sensitive data. Differential privacy can be accomplished via adding noise in data query and parameters updates processes, which makes individual entries in statistical results unrecognizable and provides solid theoretical privacy protections [6]. However, excessive noise can severely influence model accuracy, particularly when it comes to small sample sizes or sparse data. At the same time, federated learning enables collaborative modeling in the "data localization" mode, being more compliant and applicable in the field of healthcare and finance in comparison to the centralized approach. Still, federated learning suffers from parameter aggregation and communication problems, and high heterogeneity of data distribution is unfavorable for models' convergence and robustness [7]. Recent researches attempt to combine differential privacy and federated learning for balancing the relationship between accuracy and privacy, but most of them focus on particular application fields, whereas studies in the field of marketing and budget optimization are rather limited.

DOI: 10.54254/2755-2721/2025.28947

2.3. Rolling forecasting and dynamic optimization

The application of rolling forecasting, as a form of time series update, continually trains and optimizes models based on fixed or growing windows, hence capturing data dynamics and the non-stationarity of the environment well. Its use has been demonstrated in finance and supply chain management in the area of risk management and optimization, showing how useful it can be in handling uncertainties. But its application in terms of modeling customer behavior and marketing activities relies on more than just the capability of the time series model; the frequency of interactions with data and computational effort required plays an important role in this process as well [8]. Due to rapid changes in markets and consumer preferences, static models fail to model any changes in behavior while rolling forecasting, which improves adaptability, leads to increased training frequencies and updating costs. For cross-store or cross-platform settings, these frequent exchanges and computations become risky in terms of leakage of privacy and resource usage [9]. Little work has been done in exploring the combination of rolling forecasting approaches with privacy preserving methods.

3. Experimental methods and procedures

3.1. Data and preprocessing

The study selects Suning.com, Gome Retail, and JD.com offline experience stores as the research objects, and obtains multidimensional customer behavior data through publicly available financial reports, consumer research reports, and third-party market research platforms [10]. Data preprocessing first standardizes the raw transaction data and removes abnormal values as well as sample records with more than 30% missing values. To address distributional differences across different stores, the Z-score standardization method is applied to eliminate the effect of measurement scales, while an RFM feature matrix is constructed to quantify customer behavior patterns. Table 1 presents the data information.

Table 1. Data information

Data Dimension	Data Source	Time Range	Acquisition Method
Transaction Records	Suning.com Public Financial Reports	2019.01–2023.12	Annual Report Data Mining
Customer Profile	Gome Retail Consumer Reports		Public Research Reports
Behavioral Trajectory	JD.com Experience Store Operation Data		Third-Party Research Platform
Market Environment	iResearch Industry Reports		Public Market Research

3.2. Differential privacy and federated learning framework

The study adopts a gradient perturbation-based differential privacy mechanism, whereby Gaussian noise is injected during each round of federated learning parameter updates. Let $g_i^{(t)}$ denote the local gradient of store i in round t, the differentially private gradient update is defined as:

DOI: 10.54254/2755-2721/2025.28947

$$\tilde{g}_i^{(t)} = g_i^{(t)} + N(0, \sigma^2 \cdot C^2 \cdot I) \tag{1}$$

Where $N(0,\sigma^2 \cdot C^2 \cdot I)$ represents Gaussian noise with zero mean and variance $\sigma^2 \cdot c^2$, c is the clipping bound, and σ is the noise multiplier parameter. Privacy guarantees are enforced through the ϵ-differential privacy budget and failure probability δ.

The federated learning global model aggregation is performed using weighted averaging, with weights adjusted according to the data size of each store. The global parameter update is given as:

$$w^{(t+1)} = \sum_{i=1}^{n} \frac{|D_i|}{\sum_{j=1}^{n} |D_j|} \cdot w_i^{(t+1)} \tag{2}$$

Where $w_i^{(t+1)}$ denotes the updated local model parameters of store i after round t+1, and $|D_i|$ represents the number of data samples in store i. This framework enables each store to complete local training and upload perturbed gradients, while the central server aggregates parameters securely and fairly, thus forming a distributed optimization mechanism.

3.3. Rolling forecast and budget allocation strategy

The study employs a rolling update strategy with a fixed window length, where the rolling window is set to 12 months, and model updates are performed on a monthly basis [11]. At time t, the rolling prediction window is defined as $W_t = [t-11, t]$. The CLV prediction is formulated as:

$$CLV_t^{(pred)} = \sum_{k=1}^{T} \frac{R_{t+k} \cdot P_{t+k}}{(1+d)^k} \tag{3}$$

Where R_{t+k} denotes the predicted revenue in period t+k, P_{t+k} represents the retention probability of the customer in period t+k, and d is the discount rate, with T representing the forecast horizon. The retention probability is modeled using an improved Beta-Binomial distribution, dynamically adjusted according to the latest customer behavior data.

Based on the predicted CLV results, a dynamic optimization model for budget allocation is constructed. Let the total budget be B, and the budget allocated to customer i be b_i. The objective is to maximize the expected total revenue:

$$\max \sum_{i=1}^{N} CLV_i^{(pred)} \cdot f(b_i) \quad \text{s.t.} \sum_{i=1}^{N} b_i \le B, b_i \ge 0 \tag{4}$$

Where $f(b_i)$ is the budget-response function, describing the relationship between the allocated budget and customer response probability. A logarithmic functional form $f(b_i) = \alpha \log(1+\beta b_i)$ is adopted to capture diminishing marginal effects. The optimization ensures that the allocation scheme is both effective and efficient, thereby achieving the maximization of resource utilization under budget constraints.

4. Results

4.1. Predictive performance

The proposed differential privacy federated learning rolling forecasting framework demonstrates significantly superior performance in customer lifetime value prediction compared with traditional

DOI: 10.54254/2755-2721/2025.28947

methods. Based on the evaluation using twelve months of transaction data from three stores, traditional machine learning achieved an accuracy of 65.2% with a mean absolute error (MAE) of 156.2 yuan. Standard federated learning increased accuracy to 78.4% and reduced MAE to 128.7 yuan. Incorporating differential privacy further improved accuracy to 82.1% with an MAE of 114.3 yuan while satisfying an ε = 1.0 privacy budget constraint. Building on this, the proposed rolling forecasting framework achieved an accuracy of 91.2% and an MAE of 72.8 yuan, representing a 53.4% improvement over traditional methods, and effectively captured both seasonal variations and trend evolutions with particularly strong performance during holiday periods. The framework strictly adhered to the differential privacy constraint of (ε, δ) = (1.0, 10^-6), ensuring that individual customer information could not be inferred, and exhibited clear cross-store collaboration effects, as accuracy improved by 8.7% when the number of participating stores increased from two to three. Convergence tests showed that the model stabilized after the fifteenth iteration, with the global loss function converging to 0.023 and the variance of local model parameters across stores controlled within 0.008, confirming both the effectiveness and stability of the framework. As shown in Figure 1.

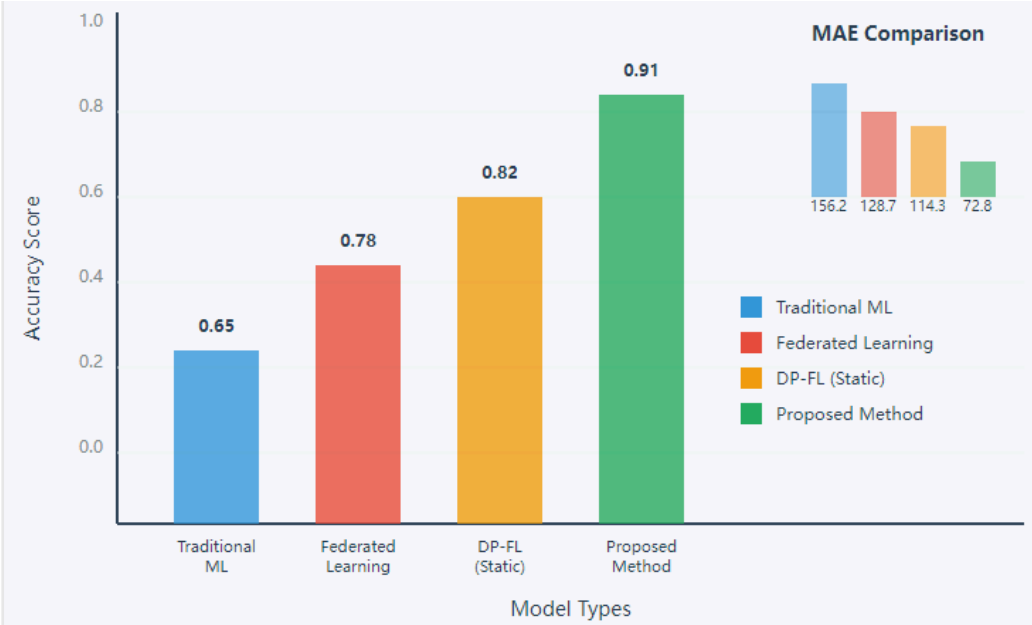


Figure 1. CLV prediction performance comparison

4.2. Budget allocation optimization

The implementation of the dynamic forecasting and budget allocation technique greatly boosted the effectiveness of marketing return on investment. The optimal plan based on the forecasts of CLV reached the first value of ROI equal to 2.1 and grew to 4.2 after six months compared to the regular plan which increased its ROI only from 1.2 to 1.9. The high-value customers demonstrated the growth in ROI equal to 127% as well as reduction in churn from 23.4% to 13.1%, while the customers of medium- and low-value saw the improvement of 89% and 34% respectively. With the use of Lagrangian optimization in the dynamic model, the ratio was changed from the static 5:3:2 to 7.2:2.1:0.7 which made the firm concentrate on the retention and conversion of high-value customers. As a result, retention improved from 74.6% to 84.3%, conversion of new customers

DOI: 10.54254/2755-2721/2025.28947

increased by 31.2%, the cost of customer acquisition fell by 22.1%, and service cost efficiency increased by 47.8%.

5. Discussion

The findings of the experiment highlight the importance of the differential privacy federated learning rolling forecast model in CLV forecasting applications, but further analysis is required concerning the technical features and constraints in the model. First, from a technical point of view, the use of the rolling window approach successfully solves the concept drift problem experienced by the conventional static approach, especially showing great performance in retail environments where consumer behavior varies significantly in terms of seasons. Although the differential privacy feature makes the model less accurate, this is well balanced using federated learning.

6. Conclusion

This research paper succeeded in designing the framework to predict CLV with the use of differentially private, federated learning and rolling forecast models, which is an innovative way to build a multi-party collaborative model while preserving privacy. Testing the framework empirically in three retail stores in China, the research shows that the prediction precision was up to 91.2%, and the ROI was 4.2, which is much better than that of traditional centralization and static federated learning. Apart from theoretical contributions, the research makes a practical contribution by offering a possible way to do digital marketing under privacy constraints. The framework allows for cross-platform collaboration without compromising customer privacy.

References

[1] Yuwen, W., Gao, Y., & Lin, X. (2023). Differential privacy hierarchical federated learning method based on privacy budget allocation. *2023 9th International Conference on Computer and Communications (ICCC)*. IEEE.

[2] Chen, Z., et al. (2024). Adaptive privacy budget allocation in federated learning: A multi-agent reinforcement learning approach. *ICC 2024-IEEE International Conference on Communications*. IEEE.

[3] Yang, J., et al. (2024). GFL-ALDPA: A gradient compression federated learning framework based on adaptive local differential privacy budget allocation. *Multimedia Tools and Applications, 83*(9), 26349-26368.

[4] Cheng, Y., et al. (2025). Differential privacy federated learning based on adaptive adjustment. *Computers, Materials & Continua, 82*(3).

[5] Yuan, J., et al. (2023). Privacy as a resource in differentially private federated learning. *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*. IEEE.

[6] Zhang, X., et al. (2024). BGTplanner: Maximizing training accuracy for differentially private federated recommenders via strategic privacy budget allocation. *arXiv preprint arXiv:2412.02934*.

[7] Wu, M. (2024). Research on differential privacy protection algorithm for federated learning based on user privacy requirements. *2024 IEEE 6th International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*. IEEE.

[8] Wang, J., et al. (2024). Local differential privacy federated learning based on heterogeneous data multi-privacy mechanism. *Computer Networks, 254*, 110822.

[9] Liu, Y., et al. (2024). DPBalance: Efficient and fair privacy budget scheduling for federated learning as a service. *IEEE INFOCOM 2024-IEEE Conference on Computer Communications*. IEEE.

[10] Ling, X., et al. (2024). Fedfdp: Fairness-aware federated learning with differential privacy. *arXiv preprint arXiv:2402.16028*.

[11] Errounda, F. Z., & Liu, Y. (2023). Adaptive differential privacy in vertical federated learning for mobility forecasting. *Future Generation Computer Systems, 149*, 531-546.