Predicting Purchase Frequency in e-Commerce: Hybrid Machine Learning Approach

Nilay İşeri, Mustafa Keskin and Onur Arda Raştak
Hepsiburada, Turkey

Keywords: e-Commerce, Purchase Frequency Prediction, Customer Behavior, Machine Learning, Multi Stage Prediction.

Abstract: This paper addresses the problem of predicting customer purchase frequency. We developed machine learning models to forecast the number of purchases a user will make next month, categorizing them into three classes. We compared multiclass classification, regression, and hybrid approaches. Our analysis shows that the most effective method is a hybrid approach that uses a binary classifier to target the 4+ purchases and a regression model for the remaining classes. This two-stage model provided a significant performance increase over single models, proving to be a robust solution for imbalanced, ordinal prediction tasks.

1 INTRODUCTION

The rapid growth of electronic commerce in recent years has led to a massive increase in online consumer activity. This surge in e-commerce usage has produced vast amounts of customer behavior data and intensified competition among online retailers. In this context, the ability to predict a customer's future purchasing behavior in particular, the number of purchases they will make in the upcoming month has become increasingly valuable. Accurate monthly purchase frequency prediction can support a range of business objectives, including personalized marketing (e.g., targeting high-frequency shoppers with loyalty rewards), inventory and supply chain optimization (forecasting product demand at the customer-segment level), and customer lifetime value (CLV) estimation (Oblander et al., 2020). Furthermore, effective prediction of individual purchasing frequency enables firms to proactively identify high-value customers and deploy retention strategies before those customers churn, thereby improving profitability (Verbeke et al., 2014).

Predicting customer purchase frequency at the individual level poses several challenges. Unlike subscription services, e-commerce customer relationships are typically *non-contractual*, meaning customers can make purchases at irregular intervals or stop purchasing at any time without notice. This irregularity leads to highly skewed purchase frequency distributions many customers make few or no purchases in a given period, while a small segment generates many orders. Traditional techniques from marketing science, such as RFM (Recency, Frequency, Monetary) scoring and probabilistic models like the Pareto/NBD and BG/NBD formulations, provide a foundation for understanding purchase frequency and customer value (Fader et al., 2005). However, these classical models rely on strong statistical assumptions (e.g., Poisson purchase timing and exponential dropout processes) and often struggle to incorporate the rich features now available (such as detailed online browsing behavior, product category preferences, and multi-channel marketing touchpoints). Moreover, in modern e-commerce datasets with millions of customers and heterogeneous behaviors, such parametric models can face scalability and accuracy limitations (Abe, 2009). These limitations have motivated a shift toward data-driven machine learning approaches, such as gradient boosting models, which can flexibly learn purchasing patterns from large-scale behavioral data without relying on predefined stochastic assumptions (Wang et al., 2023).

In light of the above, this paper focuses on developing a predictive model for monthly customer purchase frequency using boosting-based machine learning methods. The motivation for using gradient boosting is twofold: first, their proven accuracy and flexibility in similar classification/regression tasks suggest they can effectively capture the subtle factors that drive repeat purchases; second, boosting models provide feature importance metrics and explainability tools (e.g., SHAP values) that help interpret which customer attributes most strongly influence purchase

İşeri, N., Keskin, M. and Raştak, O. A.
Predicting Purchase Frequency in e-Commerce: Hybrid Machine Learning Approach.
DOI: 10.5220/0014305900004848

In *Proceedings of the 2nd International Conference on Advances in Electrical, Electronics, Energy, and Computer Sciences (ICEEECS 2025)*, pages 64-68
ISBN: 978-989-758-783-2

frequency, offering business insights in addition to predictions.

2 RELATED WORKS

The availability of large-scale behavioral data and advances in computational techniques have enabled machine learning methods such as gradient boosting to model user purchase behavior without relying on restrictive parametric assumptions (Wang et al., 2023). Algorithms such as logistic regression, random forests, and gradient boosting have been applied to predict outcomes ranging from repurchase propensity to purchase frequency or spending.

Ensemble methods, particularly gradient boosting trees, have emerged as the most effective for these tasks. Song and Liu demonstrated that XGBoost improves purchase prediction in e-commerce by incorporating diverse behavioral features (Song and Liu, 2020).Yang *et al.* extended this direction by combining Random Forest and LightGBM in a hybrid ensemble to address class imbalance and enhance repurchase prediction (Yang et al., 2021). These studies show that ensemble approaches can surpass traditional RFM and probabilistic models by leveraging richer predictors and nonlinear interactions. Feature engineering remains central, with extensions of RFM to include variables such as customer tenure, interpurchase intervals, and engagement metrics further improving predictive power. Overall, machine learning, especially ensemble learning, has become a cornerstone in purchase frequency prediction.

Wang *et al.* proposed a user purchase behavior prediction model based on XGBoost, leveraging multi-dimensional behavioral features such as historical transaction patterns, account activity metrics, and user segmentation tags to accurately forecast future purchasing behavior (Wang et al., 2023). Building on this direction, Sun *et al.* applied gradient boosting decision trees (GBDT) and random forests to predict key components of customer lifetime value (CLV), particularly purchase frequency, and showed that these models outperformed classical probabilistic approaches such as Pareto/NBD and Pareto/GGG on real-world retail datasets (Sun et al., 2021). In practical settings, CLV prediction is often decomposed into sub-tasks, churn probability, expected frequency, and average value, with boosting applied to each.

Overall, gradient boosting methods stand out for their accuracy, flexibility, and ability to incorporate diverse features in frequency prediction. They consistently outperform traditional models, though challenges remain in interpretability and integration with newer techniques.

While gradient boosting dominates structured retail data, newer methods are emerging to capture sequential dynamics and improve interpretability. Deep learning models, such as LSTMs and transformers, have been applied to customer transaction histories, treating them as sequences of events. For example, attention-based LSTMs have been used to predict high-value customer behavior, while transformer architectures have shown advantages when long and complex purchase cycles are involved (Lathwal and Batra, 2024; Kim et al., 2023).

Hybrid approaches combine machine learning with probabilistic or domain-specific modeling to balance flexibility and structure. Examples include two-stage models where neural networks predict distributional parameters for purchase counts, later refined with boosting, or reinforcement learning frameworks that go beyond prediction to optimize marketing interventions.

In sum, recent work explores deep learning for sequential behavior, hybrid models to integrate prior knowledge, and XAI methods for interpretability. Yet ensemble trees, particularly boosting algorithms, remain the strongest baseline for purchase frequency prediction, balancing accuracy, scalability, and transparency (Grinsztajn et al., 2022). These developments provide the foundation for our boosting-based framework for next-month purchase frequency prediction.

3 DATA AND PREPROCESSING

3.1 Dataset Description

The dataset was constructed from a one-year transactional history preceding the prediction month. The target was defined as the purchase frequency in June 2025, categorized into three discrete classes: (i) one purchase, (ii) two to three purchases, and (iii) four or more purchases (heavy buyers). The cohort included only active customers with at least one completed order in the historical period. The final dataset contained several million rows, with class distributions reflecting the natural skew typically observed in e-commerce frequency prediction tasks.

3.2 Feature Groups and Selection

Features originated from a customer-level datamart that aggregates behavioral, transactional, and demographic signals, accumulating nearly 600 raw attributes per member. For clarity, these features can be described in the following categories:

- **Historical behavior:** long-term purchase frequency, cumulative monetary value, product diversity.
- **Short-term recency and intensity:** recency of last purchase, recent order counts, temporal trend flags.
- **Monetary indicators:** average basket size, cumulative spending, return/refund ratios.
- **Demographic and geographic signals:** customer location identifiers (city, district, postal code).
- **Channel and payment preferences:** order origin (e.g., app vs. web), payment method, issuing bank.
- **Operational and logistic features:** address stability, historical claim activity, fraud-related markers.

While this comprehensive feature space captured diverse aspects of customer behavior, it also introduced redundancy and noise. To address this, a multi-stage feature selection pipeline was applied:

1. **Filter methods:** removal of constant or low-variance features and those with excessive missingness.
2. **Correlation analysis:** elimination of highly collinear attributes.
3. **Model-based ranking:** importance ranking using tree-based methods (e.g., LightGBM gain and split metrics).
4. **Iterative pruning:** evaluation of reduced subsets; only predictors improving stability were retained.

Through this systematic process, the original ∼600 attributes were reduced to a final set of about 40 highly informative predictors, ensuring both model efficiency and interpretability.

3.3 Preprocessing

Preprocessing steps included:

- **Missing value imputation**: mode filling for categorical features and domain-specific rules for numerics.
- **Data type normalization**: categorical identifiers (e.g., district/postal codes) restored from float to categorical.
- **Encoding**: categorical features encoded using label encoding, ensuring consistency and avoiding leakage.
- **Scaling**: continuous features standardized to zero mean and unit variance.
- **Chronological split**: training/validation sets split by time, ensuring temporally consistent evaluation.

These procedures ensured a robust, representative dataset for subsequent modeling.

4 METHODOLOGY

The goal of this study is to predict the number of products a user will purchase in the upcoming month, categorized into three classes: 1, 2–3, and 4+. We investigated multiple modeling strategies to address this problem. Two different modeling strategies were investigated.

4.1 Direct Classification

Standard supervised classification models were trained directly on the three classes. Models such as CatBoost, LightGBM, XGBoost, Extra Trees, and Logistic Regression were evaluated.

4.2 Regression with Post-Processing

As an alternative, the task was reformulated as a regression problem, where models predict a continuous estimate of the expected order count. The predicted values were then mapped into the predefined intervals (1, 2–3, 4+) to obtain class labels. This approach provides a more natural formulation of the task, since the number of orders is inherently a count variable. By treating the problem as regression, the model captures the magnitude of the outcome and preserves the ordinal structure of the classes, whereas direct classification ignores the ordering among categories. Our results demonstrate that regression-based models achieve performance comparable to, and in some cases superior to, classification models in terms of precision, recall, and F1 score, making regression a more advantageous approach in this context.

Table 1: Model performance comparison for order prediction task.

Algorithm	Accuracy	Precision	Recall	F1
XGBoost_Reg	0.590	0.608	0.522	0.539
LightGBM_Reg	0.590	0.607	0.522	0.539
CatBoost_Reg	0.591	0.607	0.520	0.538
Linear_Reg	0.580	0.612	0.512	0.527
CatBoost_Clf	0.628	0.569	0.519	0.524
LightGBM_Clf	0.628	0.570	0.518	0.523
XGBoost_Clf	0.628	0.569	0.517	0.523
ExtraTrees_Reg	0.573	0.614	0.500	0.514
ExtraTrees_Clf	0.618	0.556	0.492	0.490
Logistic_Clf	0.604	0.509	0.482	0.445

Table 2: Class based baseline model performance.

Class	Precision	Recall
1 (single purchase)	0.70	0.83
2–3 (mid-frequency)	0.46	0.44
4+ (heavy buyers)	0.55	0.29

Table 1 presents the performance comparison of different regression and classification models for the order prediction task. Overall, classification models achieve slightly higher accuracy than regression-based models, with CatBoost, LightGBM, and XGBoost classifiers all reaching an accuracy of 0.628, compared to around 0.59 for the best regression models. However, the improvement in accuracy does not translate into substantial gains in other metrics.

Regression models, particularly Linear Regression and Extra Trees Regression, demonstrate higher precision (up to 0.614) compared to classification models. This indicates that regression tends to produce fewer false positives when mapping predictions to classes. On the other hand, classification models achieve slightly better balance in terms of recall and F1 score, though none of the models surpass 0.54 in F1.

Among classifiers, CatBoost, LightGBM, and XGBoost perform similarly and represent the strongest baselines. Logistic Regression, while simpler, underperforms especially in F1 score (0.445). For regression, CatBoost Regression and LightGBM Regression yield the best overall balance of metrics.

Taken together, the results suggest that while classification yields marginally higher accuracy, regression offers better precision and leverages the ordinal structure of the problem. However, all baseline models struggled to reliably capture the 4+ class, which highlighted the need for a targeted binary formulation. As shown in Table 2, even the best-performing baseline models exhibit limited recall for this segment. This observation motivated the design of the proposed hybrid approach, which combines regression with binary classification to improve robustness across all classes.

4.3 Binary Classification

Since the initial models struggled to correctly identify users in the 4+ category, we reformulated the task as a binary classification problem: predicting whether a user will place 4+ orders or not.

Table 3 summarizes the performance of binary classifiers trained to detect whether a user belongs to the 4+ category. Gradient boosting models (LightGBM, XGBoost, and CatBoost) achieve the best overall balance, with accuracies of 0.882, AUC values around 0.85, and F1 scores near 0.45. Although recall remains relatively low (≈0.34), these models demonstrate higher precision, indicating that when they predict a user as 4+, it is often correct.

Table 3: Binary classifier performance for 4+ category prediction.

Model	Acc	AUC	Prec	Rec	F1
LightGBM_Binary	0.882	0.850	0.683	0.342	0.456
XGBoost_Binary	0.882	0.850	0.685	0.342	0.456
CatBoost_Binary	0.882	0.849	0.689	0.336	0.451
Logistic_Binary	0.875	0.803	0.651	0.291	0.403
ExtraTrees_Binary	0.868	0.832	0.843	0.106	0.189

Logistic Regression performs slightly worse, with reduced AUC and F1. Extra Trees yields the highest precision (0.843) but suffers from extremely low recall (0.106), meaning it correctly identifies very few of the actual 4+ users.

Overall, boosting-based methods provide the most reliable trade-off, though the results also highlight the inherent difficulty of predicting the 4+ class due to its limited representation in the dataset.

4.4 Hybrid Method

To further improve performance, we introduced a hybrid methodology that combines binary classification with regression. In this setup, a binary classifier (LightGBM_Binary) is first used to predict whether a user belongs to the 4+ category. If the prediction is negative, a regression model is applied to estimate the expected purchase count, which is then mapped into the 1 or 2–3 categories.

This hybrid framework leverages the strengths of both approaches: the binary classifier improves detection of the 4+ class, while regression preserves the ordinal structure of the remaining categories. Compared to baseline model(XGB_Regression), this method achieved approximately a 2% improvement across evaluation metrics, which can be seen in Table 4, demonstrating its robustness and practical value. Furthermore, class-level results in Table 5 confirm that the hybrid approach substantially increases recall for the 4+ segment while maintaining balanced performance on the other classes.

Table 4: Performance comparison of hybrid approach (LightGBM_Binary + XGB_Regression) against baseline.

Method	Accuracy	Precision	Recall	F1
Baseline	0.628	0.570	0.518	0.523
Hybrid	0.641	0.582	0.529	0.534

Figure 1 presents the two-stage framework that we developed.

Table 5: Class based hybrid model performance.

Class	Precision	Recall
1 (single purchase)	0.72	0.80
2–3 (mid-frequency)	0.48	0.46
4+ (heavy buyers)	0.55	0.33

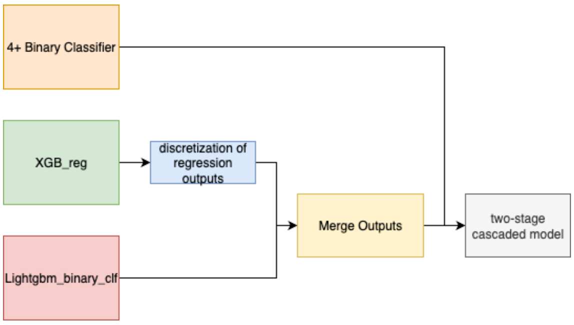

Figure 1: Two Stage Cascade Model.

5 CONCLUSIONS AND FUTURE WORKS

In this study, we addressed the problem of predicting the number of products a user will purchase in the upcoming month. We explored multiple modeling strategies, including direct classification, regression with interval mapping, and a hybrid approach combining binary classification for the 4+ category with regression for the remaining classes.

Our results indicate that while classification models achieve slightly higher overall accuracy, regression captures the ordinal and count-based nature of the target variable, resulting in better precision and more meaningful predictions. The proposed hybrid methodology successfully balances the strengths of both approaches, improving detection of the 4+ class while retaining robust performance across all categories. This demonstrates the effectiveness of combining regression and binary classification for class-imbalanced, ordinal prediction tasks in a real-world e-commerce context.

For future work, we plan to explore modern attention-based tabular learning algorithms, such as TabNet, TabTransformer, and FT-Transformer, to further improve prediction accuracy and interpretability. These models have shown strong performance on structured data by leveraging self-attention mechanisms to capture complex feature interactions. Additionally, we aim to investigate temporal and sequential patterns in user purchasing behavior, incorporating recurrent or transformer-based architectures to model dynamics over time. Combining these modern tabular methods with our hybrid framework may further enhance predictive performance, especially for 4+ category.

ACKNOWLEDGEMENTS

This project was made possible by the individual contributions of each member of the recommendation team within Hepsiburada technology group. Also, this project would not have been possible if the technology group management of Hepsiburada had not supported and encouraged the data science team in innovation.

REFERENCES

Abe, M. (2009). Counting your customers one by one: A hierarchical bayes extension to the pareto/nbd model. *Marketing Science*, 28(3):541–553.

Fader, P. S., Hardie, B. G. S., and Lee, K. L. (2005). Counting your customers the easy way: An alternative to the pareto/nbd model. *Marketing Science*, 24(2):275–284.

Grinsztajn, L., Oyallon, E., and Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on tabular data? In *Advances in Neural Information Processing Systems (NeurIPS)*. Available: https://arxiv.org/abs/2207.08815.

Kim, Y., Lee, S., and Jang, H. (2023). Customer lifetime value prediction using deep learning: A transformer-based approach. *Applied Artificial Intelligence*, 37(2):147–163.

Lathwal, P. and Batra, R. (2024). Attention-based customer lifetime value prediction in e-commerce using ft-transformer architecture. *Amity Journal of Data and Cyber Sciences*. Available: https://www.amity.edu/gurugram/jccc/pdf/JDCS_0205.pdf.

Oblander, E. S., Gupta, S., Mela, C. F., Winer, R. S., and Lehmann, D. R. (2020). The past, present, and future of customer management. *Marketing Letters*, 31(2–3):125–136.

Song, P. and Liu, Y. (2020). An xgboost algorithm for predicting purchasing behaviour on e-commerce platforms. *Tehnički vjesnik*, 27(5):1467–1471.

Sun, Y., Cheng, D., Bandyopadhyay, S., and Xue, W. (2021). Profitable retail customer identification based on a combined prediction strategy of customer lifetime value. *Midwest Social Sciences Journal*, 24(1).

Verbeke, W., Martens, D., and Baesens, B. (2014). Social network analysis for customer churn prediction. *Applied Soft Computing*, 14:431–446.

Wang, W., Xiong, W., Wang, J., Tao, L., Li, S., Yi, Y., and Zou, X. (2023). A user purchase behavior prediction method based on xgboost. *Electronics*, 12(9):2047.

Yang, L., Niu, X., and Wu, J. (2021). Rf-lightgbm: A probabilistic ensemble way to predict customer repurchase behaviour in community e-commerce. *arXiv preprint arXiv:2109.00724*.