Product Category Level Demand Forecasting Based on LRFM–CLV Customer Segmentation in B2B E-Marketplace Systems

Ariwiati
Doctoral Program of Information Systems, Postgraduate School, Universitas Diponegoro, Semarang, Indonesia
ariwiati@students.undip.ac.id (corresponding author)

Aries Susanty
Doctoral Program of Information Systems, Postgraduate School, Universitas Diponegoro, Semarang, Indonesia
aries.susanty@lecturer.undip.ac.id

Kardison Lumban Batu
Doctoral Program of Information Systems, Postgraduate School, Universitas Diponegoro, Semarang, Indonesia
kardison@lecturer.undip.ac.id

Received: 14 February 2026 | Revised: 20 March 2026 and 4 April 2026 | Accepted: 5 April 2026

 | DOI: https://doi.org/10.48084/etasr.18169

ABSTRACT

Accurate demand forecasting poses a major challenge in B2B e-marketplace systems due to the heterogeneous, complex, and nonlinear nature of transaction data. This study introduces a customer-segmentation-based demand-forecasting method that employs tailored machine learning models for each cluster. Customers were first grouped by transaction behavior characteristics, and then several prediction models, namely Random Forest, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and an averaging ensemble of these models, were evaluated within each cluster using metrics such as Mean Absolute Error (MAE), Symmetric Mean Absolute Percentage Error (SMAPE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). The results showed that no single model outperformed the others across all clusters. LightGBM yielded the best results in most clusters, with R^2 values of 0.8202 in Cluster 0, 0.7505 in Cluster 1, and 0.6816 in Cluster 2, whereas the averaging ensemble model excelled in Cluster 3 with an R^2 of 0.9849. Visualization of positively growing product categories highlighted notable differences in demand patterns between clusters. These insights suggest that segmentation-based and contextual modeling approaches can enhance both the accuracy and interpretability of demand predictions, improving support for operational decisions and supply chain planning in B2B e-marketplaces.

Keywords-customer segmentation; demand forecast; machine learning; B2B; transaction pattern analysis

I. INTRODUCTION

The rapid growth of B2B e-marketplaces has increased demand management and supply chain planning complexity due to the heterogeneous nature of B2B transactions, which are based on long-term relationships and involve large transaction volumes and values [1]. In this context, demand forecasting at the product category level is essential since product categories serve as the main decision-making units for operational planning, which includes inventory management, procurement, and distribution activities [2].

Inaccurate forecasts at this stage can cause stock imbalances, higher operating costs, and reduced service quality. Many demand forecasting methods continue to depend on aggregated historical data and overlook variations in customer behavior [3]. Recent studies indicate that demand forecasting using data mining and machine learning at the product category level can enhance prediction accuracy by capturing more detailed temporal patterns and demand traits [4-6]. Nonetheless, by ignoring customer heterogeneity, forecasting models often produce average estimates that do not accurately reflect true demand fluctuations in a B2B setting.

To address these limitations, transaction-based customer segmentation is commonly used in B2B analytics systems. The Length, Recency, Frequency, Monetary (LRFM) model effectively classifies customers based on the duration of their relationship, the recency of their transactions, the frequency of their purchases, and the value of their transactions [7-9]. At the same time, Customer Lifetime Value (CLV) offers insight into the long-term financial worth of customers. Recent studies indicate that combining LRFM and CLV yields more strategic and relevant customer segmentation, thereby improving operational and marketing decisions on digital B2B platforms [10, 11].

However, incorporating LRFM–CLV segmentation into demand forecasting models remains limited. Most research treats customer segmentation as an independent outcome rather than as a factor within forecasting models. Consequently, variations in demand across segments, particularly at the product category level, are not adequately captured in predictions. Recent studies highlight that using customer behavior data in predictive models can enhance accuracy and reinforce decision-making support [12]. Consistent with these findings, another study indicates that using predictive analytics in strategic decision-making significantly enhances organizational adaptability, profitability, and sustainability, particularly for small businesses in a dynamic and competitive environment [13].

Building on this gap, the study introduces a demand-forecasting method at the product category level that utilizes LRFM–CLV customer segmentation within a B2B e-marketplace system. Customer segmentation provides a framework for modeling demand patterns within categories with greater precision, based on customer behavior. This strategy aims to enhance forecasting accuracy and offer better support for inventory planning and supply chain decisions in a B2B e-marketplace setting.

This study employs several widely used machine learning models for demand forecasting and predictive analytics, including Random Forest, Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM). Additionally, an ensemble approach was created to leverage the strengths of each model, enhancing prediction stability and accuracy. The performance of each model was evaluated and compared to determine the optimal configuration for demand forecasting at the product category level, based on LRFM–CLV customer segmentation.

II. METHODOLOGY

This study employed a quantitative approach, data mining, and machine learning methods to forecast demand at the product category level within a B2B e-marketplace system. The methodology used LRFM–CLV-based customer segmentation to analyze customer behavior in demand forecasting. As illustrated in Figure 1, the overall research process involves data collection and preprocessing, customer segmentation, demand forecasting model development, and model performance assessment.

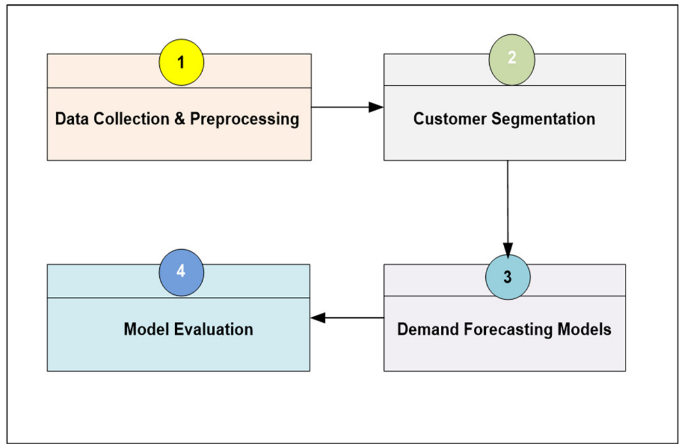


Fig. 1. Research methodology stages.

A. *Data Collection and Preprocessing*

This study utilized a dataset derived from historical transaction records collected from a B2B e-marketplace platform through internal database access. The dataset consists of 530,049 transaction records with 26 attributes, including transaction time, customer information, product details, product category, geographic location, and transaction value. To illustrate the structure of the dataset, Table I presents a small illustrative excerpt of the transaction records. It should be noted that this table only shows a subset of the data for readability purposes.

TABLE I. SAMPLE DATASET OF B2B E-MARKETPLACE PLATFORM TRANSACTION HISTORY

Order date	Customer group ID	Customer segment	Main product category	Subcategory	Quantity	Unit price (USD)
2024-04-08	631e5db2516d7c2adcebc50c	3	Food & beverages	Snack	1	0.88
2024-04-11	631e660d516d7c2adcec0cf1	0	Electronic	Audio	4	44.12
2024-04-08	6320490336adb3394d50719f	0	Handicraft	Building materials	2	802.94
2024-04-08	6320490336adb3394d50719f	0	Handicraft	Machinery	2	20.06
2024-04-09	631e5ddf516d7c2adcebc6ad	3	Food & beverages	Food packages and parcels	4	0.76
2024-04-09	63207d6036adb3394d51e5c6	3	Food & beverages	Dairy products	1	1.18
2024-04-12	631e6898516d7c2adcec27a1	3	Food & beverages	Drinks	1	1.74
2024-04-13	631e6f19516d7c2adcec6959	3	Food & beverages	Drinks	1	0.76

The dataset was obtained from a B2B e-marketplace platform operating in the electronics and consumer goods sector in Indonesia. The data cover the period from January 2023 to December 2024. Data extraction was performed through authorized internal database access, and all customer-related identifiers were anonymized to ensure privacy protection. No personally identifiable information was included in the dataset.

The temporal attributes of the dataset include order_date, year, month, week, and quarter, which define the time dimension used for demand forecasting. Customer-related data are represented by buyer_group_id and customer segmentation results, reflecting variations in customer behavior. Product-related attributes include product_id, product_name, brand, and hierarchical product categories (main_cat, sub_cat1, sub_cat2, and sub_cat3). In addition, quantitative attributes such as quantity and price_per_item are used to represent transaction volume and value.

Prior to modeling, data preprocessing was performed to improve data quality and consistency, including handling missing and inconsistent values, standardizing temporal attributes, and validating numerical variables such as quantity and price. These steps follow common practices in data mining and transaction analytics as reported in previous studies [14]. Furthermore, customer segmentation results obtained from clustering were integrated as contextual variables in the forecasting model.

B. Customer Segmentation Using Length, Recency, Frequency, and Monetary and Customer Lifetime Value

This study's customer segmentation aimed to capture the diversity of B2B customer transaction behaviors that were not fully represented by aggregate data. The method combined the LRFM model with CLV to thoroughly describe customer behavior and economic worth, leveraging common practices in transaction data-based segmentation and customer value analysis [15, 16].

The LRFM variable measured the time and financial aspects of customer relationships, including relationship length (Length), time since last transaction (Recency), transaction frequency (Frequency), and transaction monetary value (Monetary); these metrics have proven effective in distinguishing customer behavior patterns across various digital business settings, including B2B systems. Additionally, CLV estimated the long-term value that customers brought to the business, ensuring that segmentation considered not only past transaction history but also the strategic importance of customers to the company. To ensure reproducibility and clarity, the LRFM variables are formally defined as follows:

- Length (L) represents the duration of the customer relationship, calculated based on (1):

  $$L = t_{last} - t_{first} \tag{1}$$

  where t_{first} and t_{last} denote the first and last transaction times.

- Recency (R) represents the time since the last transaction and can be calculated based on (2):

  $$R = t_{ref} - t_{last} \tag{2}$$

  where t_{ref} is the reference time.

- Frequency (F) represents the total number of transactions and can be calculated using (3):

  $$F = \sum_{i=1}^{n} 1 \tag{3}$$

  where n is the total number of transactions associated with a customer, and i is the index of each transaction. The summation of 1 for each transaction effectively counts the total number of transactions.

- Monetary (M) represents the total value of transactions and can be calculated using (4):

  $$M = \sum_{i=1}^{n} v_i \tag{4}$$

  where v_i is the value of each transaction.

- CLV (CLV) is calculated using (5):

  $$CLV = \frac{M \times F}{R} \tag{5}$$

Customers were grouped using the K-Medoids clustering algorithm based on LRFM and CLV values. This method was selected for its robustness against outliers and its effectiveness in managing asymmetric transaction data distributions. The clustering process was implemented using the Euclidean distance metric to measure similarity between customer profiles. The optimal number of clusters (k) was determined using the Silhouette Score method, which evaluates both cluster cohesion and separation. Based on this analysis, the optimal number of clusters was identified as $k = 4$.

To ensure the robustness of the clustering results, the quality of the clusters was assessed using silhouette coefficient values and distribution consistency across segments. The evaluation results indicate that the clusters are well-separated and exhibit distinct behavioral characteristics.

The customer segmentation results were then remapped onto transaction data and used as contextual variables in demand-forecasting models at the product category level. Figure 2 shows the distribution of customers across all LRFM–CLV-based segments identified through the K-Medoids clustering algorithm. The distribution is uneven, with Segment 3 being the largest, followed by Segments 0 and 2, whereas Segment 1 has the fewest customers. This pattern underscores the diversity of customer behaviors in the B2B electronics market and underscores the importance of customer segmentation as contextual data for category-level demand forecasting.

Figure 3 illustrates the normalized profiles of the variables LRFM and CLV for each customer segment identified through K-Medoids clustering. Normalization was applied to facilitate direct comparisons across segments, thereby facilitating the observation of variations in customer behavior and economic value.

The visualization results show that each customer segment has distinct characteristics. Segment 0 is characterized by relatively high values for Length, Frequency, and Monetary, and low Recency, indicating active customers with long-term relationships who contribute significantly to transactions. In contrast, Segment 2 shows high Recency and low values for the other variables, indicating passive customers or those at risk of churn.

Segment 1 has the highest CLV despite relatively low transaction frequency, indicating that customers with high long-term economic value do not transact frequently. Meanwhile, Segment 3 is characterized by a fairly long relationship duration, low transaction contributions, and low economic value, reflecting regular customers with small but significant transaction values.

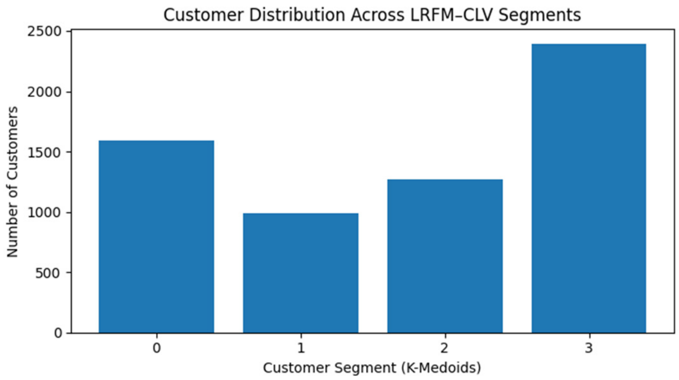


Fig. 2. Distribution of customers across LRFM–CLV segments using K-Medoids clustering.

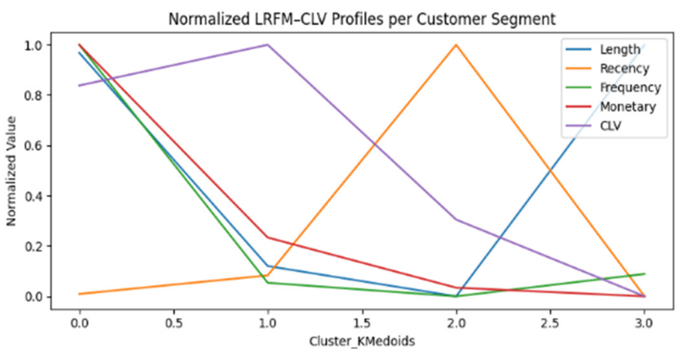


Fig. 3. Normalized LRFM–CLV profiles per customer segmentation.

C. Demand Forecasting Models

In this work, demand is defined as the total quantity of products per category within a given time period. The transaction data were aggregated on a weekly basis to construct the demand time series, with a forecasting horizon set to one week ahead.

To prepare the data for modeling, a time-based splitting strategy was applied, where earlier observations were used for training and more recent observations were reserved for testing. This approach reflects real-world forecasting scenarios and ensures that no future information is used during model training. To further prevent information leakage, all feature engineering processes were performed using only past data relative to the prediction time.

Machine learning-based demand forecasting methods are widely applied in business, e-commerce, and supply chain contexts to support operational decision-making and adaptive resource management [17, 18]. The demand forecasting in this study was conducted at the product category level, with customer segmentation used as a contextual variable. Transaction data, enriched with customer segment details, were transformed into forecasting features specific to each product category. These features included temporal attributes such as year, month, week, and quarter, as well as the product category structure and customer segment information, all treated as categorical variables. Additionally, numeric features such as the number of transactions and the total demand value per period were incorporated to capture historical demand trends. This feature set enables the model to learn nonlinear relationships among time, product categories, and customer segments that shape demand patterns.

To ensure a fair and comprehensive evaluation, both baseline forecasting methods and machine learning models were considered in this study. Simple baseline approaches, including Naive Forecasting and Moving Average, were first implemented to provide reference performance benchmarks.

In addition, four machine learning-based forecasting models were evaluated: Random Forest, XGBoost, LightGBM, and an ensemble approach. Random Forest was selected for its robustness in handling nonlinear relationships and high-dimensional data while reducing overfitting [19]. XGBoost and LightGBM, as gradient boosting algorithms, iteratively enhance prediction accuracy by optimizing decision tree structures [20, 21]. XGBoost is widely recognized for its effectiveness in modeling complex tabular data [22], whereas LightGBM offers improved computational efficiency and scalability for large-scale datasets [23]. An ensemble forecasting strategy was further implemented by combining predictions from Random Forest, XGBoost, and LightGBM using simple averaging. This approach aims to improve prediction stability and generalization performance by leveraging the complementary strengths of individual models.

D. Model Evaluation

The performance of the demand forecasting models was evaluated to assess their ability to predict product-category demand in a B2B e-marketplace environment. The evaluation focused on Random Forest, XGBoost, LightGBM, and the ensemble approach. A time-based evaluation strategy was applied, where models were trained on historical data and tested on more recent observations to simulate real-world forecasting conditions and prevent information leakage. Model performance was measured using four widely adopted metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Symmetric Mean Absolute Percentage Error (SMAPE), and the coefficient of determination (R^2), defined as follows:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (7)$$

$$\text{SMAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (9)$$

where y_i represents the actual demand, $\hat{y}_i$ denotes the predicted demand, $\bar{y}$ is the mean of the actual values, and n is the number of observations. SMAPE was used instead of MAPE to provide more stable error measurements, particularly when actual values are close to zero.

III. RESULTS AND DISCUSSION

The experiments in this study were conducted across several testing scenarios to evaluate the performance of product-category demand prediction models in a customer-segmentation-based B2B e-marketplace system. This evaluation aimed to compare the performance of several decision-tree-based machine learning models and to assess the contribution of ensemble strategies to improving the accuracy and stability of predictions before their use in the demand forecasting stage.

In the initial scenario, each model, namely, Random Forest, XGBoost, and LightGBM, was trained separately on individual customer clusters. These models were evaluated during the same testing period, enabling a direct comparison of their prediction error rates and generalization ability, as measured by the employed evaluation metrics.

In the second scenario, an ensemble method was used by averaging the predictions from the three individual models. This approach aimed to determine whether combining predictions could reduce variance and yield more stable results than standalone models. All scenarios were assessed with the same metrics: MAE, SMAPE, RMSE, and R^2. This ensured that any performance differences directly reflected the different modeling strategies.

The evaluation results from each scenario were analyzed to identify performance patterns among models within each cluster and to select the best-performing model. This model was then employed in the demand forecasting stage to generate more accurate, context-specific demand predictions.

Table II presents the performance comparison of all evaluated models, including baseline approaches (Naive and Moving Average) and machine learning models (Random Forest, XGBoost, LightGBM, and ensemble), across all customer clusters using MAE, SMAPE, RMSE, and R^2 metrics. The inclusion of baseline models provides a reference point to assess whether the proposed segmentation-based machine learning approach yields meaningful improvements beyond simple forecasting techniques.

Overall, the results demonstrate that all machine learning models consistently outperform the baseline methods across all clusters. This indicates that the observed performance gains are not due to favorable experimental conditions, but rather reflect the effectiveness of incorporating LRFM–CLV-based customer segmentation into the forecasting process.

Model performance varies across clusters, highlighting the heterogeneous nature of demand patterns in the B2B e-marketplace environment. In Cluster 0, LightGBM achieves the best overall performance, with the lowest MAE and RMSE and the highest R^2, indicating its strong capability in capturing nonlinear demand patterns. Although the ensemble model slightly improves SMAPE, LightGBM provides a more balanced performance across all evaluation metrics.

In Cluster 1, all models exhibit relatively low error values, suggesting a more stable demand structure. LightGBM again demonstrates the best performance, outperforming both baseline and other machine learning models, confirming its suitability for modeling consistent demand behavior.

In Cluster 2, model performance varies significantly, reflecting a more complex and volatile demand pattern. The Random Forest model produces a negative R^2, indicating poor explanatory capability. In contrast, LightGBM significantly improves predictive performance, achieving the highest R^2 and the lowest error values among all models. Compared to the baseline methods, the improvement is substantial, demonstrating the importance of using advanced models combined with customer segmentation for handling complex demand dynamics.

Cluster 3 exhibits highly predictable demand patterns, with all machine learning models achieving R^2 values above 0.98. The ensemble model performs best in this cluster, yielding the lowest MAE and RMSE and the highest R^2. This suggests that combining multiple models enhances prediction stability in scenarios with strong and consistent demand signals.

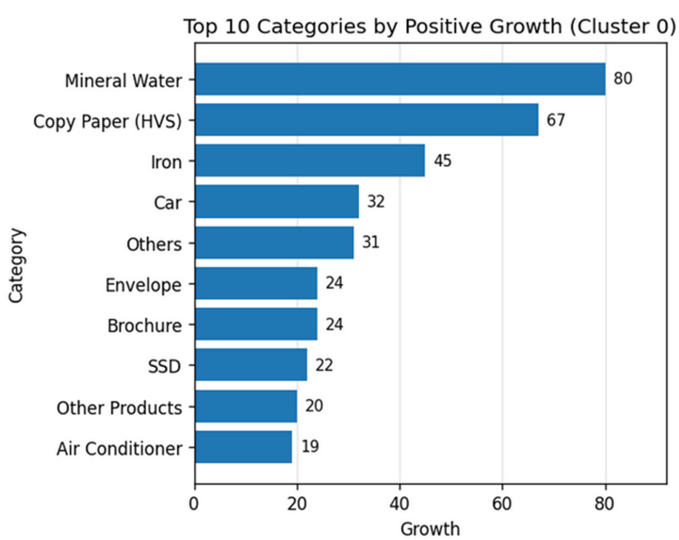
TABLE II. COMPARISON OF PREDICTION MODEL PERFORMANCE ON EACH CLUSTER

Cluster	Model	MAE	SMAPE (%)	RMSE	R^2
0	Naive	5.12	70.21	21.45	0.42
0	Moving Average	4.68	65.10	19.87	0.50
0	Random Forest	3.4017	58.94	17.4919	0.7942
0	XGBoost	3.345	58.68	17.0759	0.8039
0	LightGBM	3.1482	57.55	16.3513	0.8202
0	Ensemble	3.1864	56.98	16.506	0.8168
1	Naive	1.45	72.10	3.50	0.48
1	Moving Average	1.20	69.00	3.10	0.55
1	Random Forest	0.9854	66.60	2.5089	0.7344
1	XGBoost	0.9749	65.96	2.6395	0.706
1	LightGBM	0.9708	66.39	2.4316	0.7505
1	Ensemble	0.9697	66.44	2.4502	0.7467
2	Naive	1.60	74.50	6.20	0.10
2	Moving Average	1.30	70.00	5.10	0.30
2	Random Forest	0.9973	61.12	5.5791	-0.357
2	XGBoost	0.9092	59.44	3.959	0.3167
2	LightGBM	0.8903	56.87	2.7026	0.6816
2	Ensemble	0.884	57.91	3.2966	0.5262
3	Naive	2.00	65.00	4.50	0.80
3	Moving Average	1.60	60.00	4.00	0.88
3	Random Forest	1.2923	56.79	3.6649	0.9827
3	XGBoost	1.302	56.39	3.7195	0.9822
3	LightGBM	1.2526	56.26	3.6228	0.9831
3	Ensemble	1.2089	55.64	3.4243	0.9849

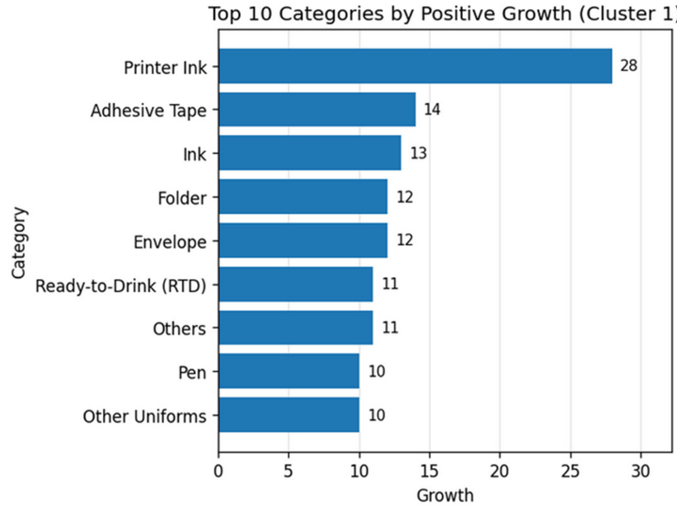
Based on these findings, the best-performing model for each cluster was selected for the forecasting stage. Forecasting was conducted separately for each cluster using historical demand data, ensuring that cluster-specific demand characteristics were preserved. A chronological data split was applied, where models were trained on past observations and evaluated on future periods, thereby maintaining the temporal integrity of the data and preventing information leakage. This approach enables the models to capture realistic demand trends and supports more reliable forecasting in practical B2B e-marketplace scenarios.

Figure 4 shows the distribution of the top 10 product categories with the highest positive growth in each customer segment, highlighting the differences in their characteristics and demand trends within the B2B e-marketplace system. For Cluster 0 and Cluster 1, growth is primarily driven by product categories tied to operational requirements and routine usage, suggesting a stable, repetitive, and predictable demand pattern. In contrast, Cluster 2 exhibits lower growth rates across categories, suggesting more specialized, less aggressive, and niche-market demand. Meanwhile, Cluster 3 is characterized by fast-moving consumer goods with steady growth, indicating high transaction volumes and persistent demand. Overall, this visualization highlights the varied product growth patterns across clusters and underscores the importance of using customer segmentation for forecasting. This approach helps generate more contextual, realistic, and relevant demand projections for operational decision-making.

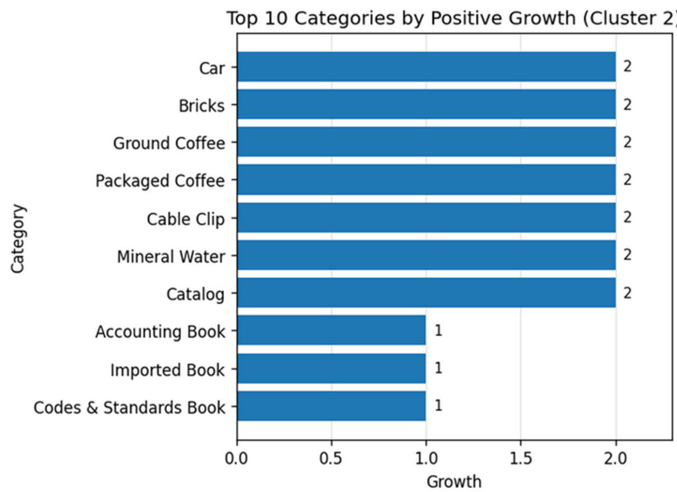
Experimental results show that combining customer segmentation with machine learning forecasting enhances understanding of demand trends in B2B e-marketplaces. Variations in product category growth across clusters suggest that demand behavior varies and is affected by distinct customer traits. This aligns with prior studies indicating that segmentation-based forecasting offers greater prediction accuracy and clarity than a single aggregate model, particularly in complex and nonlinear B2B settings. Practically, these findings significantly impact operational and managerial decisions. Customer segmentation allows companies to customize their inventory planning, procurement, and distribution based on each cluster's demand patterns. Clusters with steady, recurring demand can be handled using more predictive and efficient inventory strategies, whereas those with more variable or niche demand need a flexible and adaptable approach.



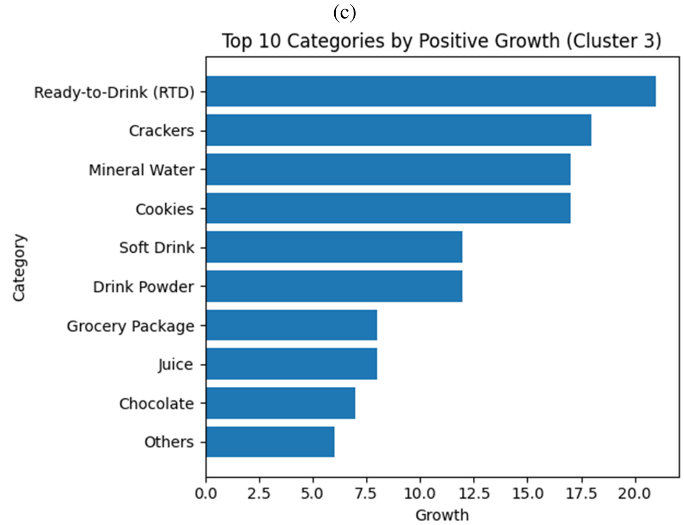
(a)



(b)



(c)



(d)

Fig. 4. Top 10 product categories with positive growth in each customer cluster: (a) Cluster 0, (b) Cluster 1, (c) Cluster 2, (d) Cluster 3.

IV. CONCLUSION

This study proposes a customer-segmentation-based demand forecasting framework that integrates Length, Recency, Frequency, Monetary (LRFM)–Customer Lifetime Value (CLV) segmentation with machine learning models for product category level prediction in B2B e-marketplace systems. The results demonstrate that incorporating customer segmentation as a contextual variable significantly improves forecasting performance compared to traditional aggregate approaches.

The evaluation shows that Light Gradient Boosting Machine (LightGBM) achieved the best performance in most clusters, with R^2 values of 0.8202, 0.7505, and 0.6816 in Clusters 0–2, whereas the averaging ensemble model outperformed others in Cluster 3 with an R^2 of 0.9849. These findings indicate that model effectiveness varies across customer segments, highlighting the importance of segment-specific modeling strategies.

The main contribution of this study lies in integrating LRFM–CLV-based customer segmentation into demand forecasting models, enabling more accurate, context-aware predictions that better capture heterogeneous transaction patterns in B2B environments. From a practical perspective, the proposed approach provides a robust foundation for operational decision-making, particularly in inventory planning and supply chain management. By leveraging segment-specific demand insights, organizations can optimize procurement and distribution strategies while reducing the risks of overstocking and understocking.

Future research may extend this work by incorporating additional contextual variables, such as seasonal effects and promotional activities, as well as exploring advanced approaches, including hybrid models and deep learning techniques, to further enhance forecasting performance and generalizability.

DECLARATION OF COMPETING INTERESTS

Not applicable to this work.

ACKNOWLEDGMENT

The authors would like to thank all parties who supported this research. We acknowledge the B2B e-marketplace platform for providing access to the anonymized transaction dataset used in this study. We also express our sincere gratitude to Universitas Diponegoro for its academic support and research environment.

DATA AVAILABILITY

The dataset used in this study was obtained from a B2B e-marketplace platform operating in the electronics and consumer goods sector in Indonesia. Due to confidentiality agreements, the data are not publicly available.

AI USE AND DECLARATION OF GENERATIVE AI USE

The authors declare that generative Artificial Intelligence (AI) tools were used in this study to assist in language editing and improve the clarity of the manuscript. The authors take full responsibility for the content of the publication.

REFERENCES

[1] D. Mourtzis, J. Angelopoulos, and N. Panopoulos, "A survey of digital B2B platforms and marketplaces for purchasing industrial product service systems: A conceptual framework," *Procedia CIRP*, vol. 97, pp. 331–336, Jan. 2021, https://doi.org/10.1016/j.procir.2020.05.246.

[2] R. M. van Steenbergen and M. R. K. Mes, "Forecasting demand profiles of new products," *Decision Support Systems*, vol. 139, Dec. 2020, Art. no. 113401, https://doi.org/10.1016/j.dss.2020.113401.

[3] A. Kumar, R. Shankar, and N. R. Aljohani, "A big data driven framework for demand-driven forecasting with effects of marketing-mix variables," *Industrial Marketing Management*, vol. 90, pp. 493–507, Oct. 2020, https://doi.org/10.1016/j.indmarman.2019.05.003.

[4] J. Feizabadi, "Machine learning demand forecasting and supply chain performance," *International Journal of Logistics Research and Applications*, vol. 25, no. 2, pp. 119–142, Feb. 2022, https://doi.org/10.1080/13675567.2020.1803246.

[5] I. H. Ariyanto, A. A. Chamid, and R. Fiati, "Demand Prediction and Apparel Production Management Using AI-Based Decision Tree," *bit-Tech*, vol. 8, no. 1, pp. 59–67, Aug. 2025, https://doi.org/10.32877/bt.v8i1.2325.

[6] B. Huang, S. Chen, and H. Chen, "Product Demand Forecasting Method Based on Spatiotemporal Hypergraph Attention Network," *Applied Sciences*, vol. 15, no. 24, Dec. 2025, Art. no. 13196, https://doi.org/10.3390/app152413196.

[7] M. R. K. Ibrahim and R. Tyasnurita, "LRFM Model Analysis for Customer Segmentation Using K-Means Clustering," in *2022 International Conference on Electrical and Information Technology*, Malang, Indonesia, 2022, pp. 383–391, https://doi.org/10.1109/IEIT56384.2022.9967896.

[8] O. N. Akande, H. B. Akande, E. O. Asani, and B. T. Dautare, "Customer Segmentation through RFM Analysis and K-means Clustering: Leveraging Data-Driven Insights for Effective Marketing Strategy," in *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals*, Omu-Aran, Nigeria, 2024, pp. 1–8, https://doi.org/10.1109/SEB4SDG60871.2024.10630052.

[9] Y. Syahra, A. Fadlil, and H. Yuliansyah, "Customer Segmentation Using RFM and K-Means Clustering to Support CRM in Retail Industry," *Sinkron : jurnal dan penelitian teknik informatika*, vol. 9, no. 3, pp. 1120–1131, July 2025, https://doi.org/10.33395/sinkron.v9i3.14907.

[10] F. A. Laksono, B. Rachmat, and Y. Sutasrso, "B2B Customer Segmentation Based on Customer Lifetime Value Concept and RFM Modeling," *International Journal of Economics Development Research*, vol. 4, no. 3, pp. 1856–1874, Dec. 2023, https://doi.org/10.37385/ijedr.v5i1.4295.

[11] Ariwiati, A. Susanty, and K. L. Batu, "An Integrated Lrfm and CLV-Based Customer Segmentation Model for B2B E-Marketplace Platform," in *2025 12th International Conference on Information Technology, Computer, and Electrical Engineering*, Semarang, Indonesia, 2025, pp. 1–6, https://doi.org/10.1109/ICITACEE66165.2025.11232828.

[12] H. GhorbanTanhaei, P. Boozary, S. Sheykhan, M. Rabiee, F. Rahmani, and I. Hosseini, "Predictive analytics in customer behavior: Anticipating trends and preferences," *Results in Control and Optimization*, vol. 17, Dec. 2024, Art. no. 100462, https://doi.org/10.1016/j.rico.2024.100462.

[13] E. Ezeife, M. E. Eyeregba, C. Mokogwu, and T. D. Olorunyomi, "Integrating predictive analytics into strategic decision-making: A model for boosting profitability and longevity in small businesses across the United States," *World Journal of Advanced Research and Reviews*, vol. 24, no. 2, pp. 2490–2507, Nov. 2024, https://doi.org/10.30574/wjarr.2024.24.2.3635.

[14] H. Ahaggach, L. Abrouk, and E. Lebon, "Systematic Mapping Study of Sales Forecasting: Methods, Trends, and Future Directions," *Forecasting*, vol. 6, no. 3, pp. 502–532, July 2024, https://doi.org/10.3390/forecast6030028.

[15] M. Husnah and R. A. Vinarti, "Customer Segmentation Analysis Using LRFM Based Product and Brand Dimensions," in *2023 2nd*

International Conference for Innovation in Technology, Bangalore, India, 2023, pp. 1–6, https://doi.org/10.1109/INOCON57975.2023.10100974.

[16] C. Nisak and E. Sugiharti, "Customer Lifetime Value Clustering Using K-Means Algorithm with Length Recency Frequency Monetary Model to Enhance Customer Relationship Management," *Journal of Advances in Information Systems and Technology*, vol. 6, no. 1, pp. 149–165, Aug. 2024, https://doi.org/10.15294/jaist.v6i1.5011.

[17] Z. Yu, "Machine Learning-Based Market Demand Forecasting in Product Development," in *2024 IEEE 6th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability*, Tainan, Taiwan, 2024, pp. 425–427, https://doi.org/10.1109/ECBIOS61468.2024.10885465.

[18] M. Zohdi, M. Rafiee, V. Kayvanfar, and A. Salamiraad, "Demand forecasting based machine learning algorithms on customer information: an applied approach," *International Journal of Information Technology*, vol. 14, no. 4, pp. 1937–1947, June 2022, https://doi.org/10.1007/s41870-022-00875-3.

[19] Z. Sun, G. Wang, P. Li, H. Wang, M. Zhang, and X. Liang, "An improved random forest based on the classification accuracy and correlation measurement of decision trees," *Expert Systems with Applications*, vol. 237, Mar. 2024, Art. no. 121549, https://doi.org/10.1016/j.eswa.2023.121549.

[20] S. Hakkal and A. A. Lahcen, "XGBoost To Enhance Learner Performance Prediction," *Computers and Education: Artificial Intelligence*, vol. 7, Dec. 2024, Art. no. 100254, https://doi.org/10.1016/j.caeai.2024.100254.

[21] D. Demirtürk, Ö. Mintemur, and A. Arslan, "Optimizing LightGBM and XGBoost Algorithms for Estimating Compressive Strength in High-Performance Concrete," *Arabian Journal for Science and Engineering*, vol. 51, no. 4, pp. 4401–4423, Feb. 2026, https://doi.org/10.1007/s13369-025-10217-7.

[22] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Information Fusion*, vol. 81, pp. 84–90, May 2022, https://doi.org/10.1016/j.inffus.2021.11.011.

[23] K. A. B. Hamou, Z. Jarir, and S. Elfirdoussi, "Application of LightGBM Algorithm in Production Scheduling Optimization on Non-Identical Parallel Machines," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 17973–17978, Dec. 2024, https://doi.org/10.48084/etasr.8779.