RESEARCH ARTICLE • OPEN ACCESS

Integrating Customer Segmentation, Predictive Modelling, and Uplift Modelling for Retail Voucher Targeting

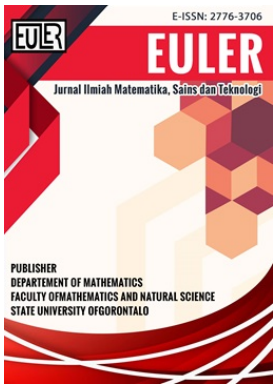
Michael Calvin Charmelino Wijaya and Maria Zefanya Sampe

Volume 14, Issue 2, Pages 522–538, August 2026

Received 19 June 2026, Revised 27 July 2026, Accepted 3 August 2026, Published 18 August 2026

To Cite this Article : M. C. C. Wijaya and M. Z. Sampe,"Integrating Customer Segmentation, Predictive Modelling, and Uplift Modelling for Retail Voucher Targeting", *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 14, no. 2, pp. 522–538, 2026, *https://doi.org/10.37905/euler.v14i2.39943*

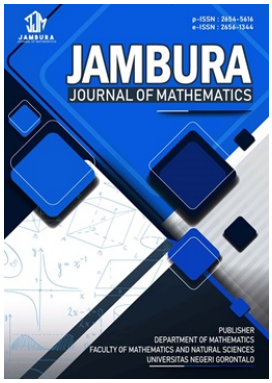
JOURNAL INFO • EULER : JURNAL ILMIAH MATEMATIKA, SAINS DAN TEKNOLOGI



Homepage	:	http://ejurnal.ung.ac.id/index.php/euler/index
Journal Abbreviation	:	Euler J. Ilm. Mat. Sains dan Teknol.
Frequency	:	Three times a year
Publication Language	:	English (preferable), Indonesia
DOI	:	https://doi.org/10.37905/euler
Online ISSN	:	2776-3706
License	:	Creative Commons Attribution-NonCommercial 4.0 International License
Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
Country	:	Indonesia
OAI Address	:	http://ejurnal.ung.ac.id/index.php/euler/oai
Google Scholar ID	:	QF_r-_gAAAAJ
Email	:	euler@ung.ac.id

JAMBURA JOURNAL • FIND OUR OTHER JOURNALS

Jambura Journal of Biomathematics



Jambura Journal of Mathematics

Jambura Journal of Mathematics Education

Jambura Journal of Probability and Statistics

Euler : Jurnal Ilmiah Matematika, Sains dan Teknologi, Volume 14, Issue 2, pp. 522–538, August 2026 *https://doi.org/10.37905/euler.v14i2.39943*

Research Article

Check for updates

Integrating Customer Segmentation, Predictive Modelling, and Uplift Modelling for Retail Voucher Targeting

Michael Calvin Charmelino Wijaya[1], **Maria Zefanya Sampe**[1,*]

[1]*Department of Mathematics, Universitas Prasetiya Mulya, Tangerang, 15339, Indonesia*

ARTICLE HISTORY
Received 19 June 2026
Revised 27 July 2026
Accepted 3 August 2026
Published 18 August 2026

KEYWORDS
K-Means Clustering
XGBoost
CatBoost
Uplift Modelling
Voucher Targeting
E-Commerce

ABSTRACT. *The rapid growth of e-commerce has encouraged retail companies to optimize data-driven promotional strategies to achieve higher targeting precision and cost efficiency. Mass promotional campaigns without proper segmentation often lead to budget inefficiencies and reduced profit margins. This study aims to optimize voucher targeting strategies for transactions involving growing-up milk products during the January–December 2024 period by integrating customer segmentation, predictive modelling, and uplift modelling. Customer segmentation was performed using K-Means Clustering, resulting in seven distinct clusters representing heterogeneous purchasing behaviors. The clustering output was incorporated as an additional feature in classification models based on XGBoost and CatBoost to predict the probability of targeted voucher redemption. Model evaluation using ROC–AUC, PR–AUC, confusion matrix, and classification report indicated that XGBoost with cluster features achieved the best performance, with a ROC–AUC of 0.9972 and a PR–AUC of 0.5479. Since conventional classification does not measure the causal impact of promotions, uplift modelling with a two-model approach was implemented to estimate the incremental effect of voucher distribution. The results reveal that approximately 1,027 customers (0.46% of 222,036 total customers) belong to the persuadable segment, generating an uplift of 1.91%. These findings demonstrate that uplift-based targeting is more efficient than mass campaigns or probability-based predictive targeting alone. The integration of segmentation, predictive modelling, and uplift modelling provides a more precise and measurable framework for data-driven promotional strategies in retail e-commerce.*

1. Introduction

Digitalization over the past decade has significantly accelerated the growth of the e-commerce sector both globally and nationally. A report by the International Trade Administration indicates that global B2B e-commerce transaction value has increased rapidly and is projected to continue growing consistently [1]. Recent data also show that Indonesia's e-commerce transaction value reached IDR 512 trillion in 2024 [2], and Indonesia is projected to become the largest e-commerce market in Southeast Asia, with its e-commerce GMV projected to reach IDR 2.39 quadrillion in 2030 [3, 4]. The development of Indonesia's e-commerce transaction value is illustrated in Figure 1.

This growth is supported by the increasing number of active e-commerce users, which is projected to continue rising until 2030 [2]. However, the contribution of e-commerce to total national retail transactions remains relatively small compared to offline retail, indicating substantial room for further growth [2]. On the other hand, competition among platforms has become increasingly intense, particularly through promotional strategies based on discounts and vouchers. The number of e-commerce users in Indonesia is shown in Figure 2.

*Corresponding Author.

Email : *maria.zefanya@prasetiyamulya.ac.id* (M. Z. Sampe)
Homepage : http://ejurnal.ung.ac.id/index.php/euler/index / E-ISSN : 2776-3706

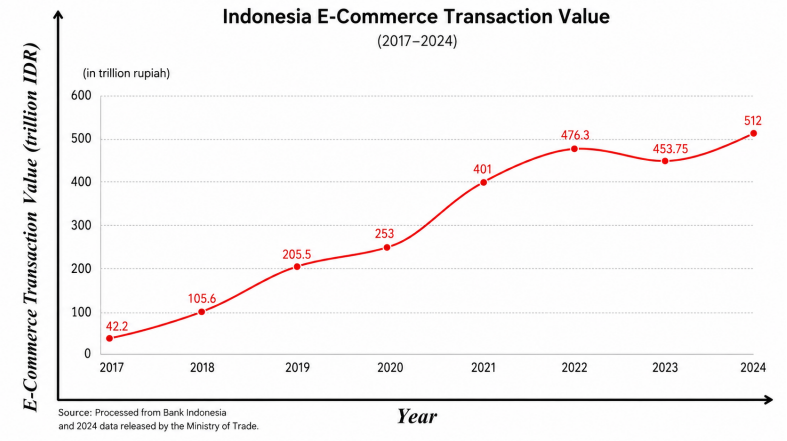


Figure 1. Indonesia e-commerce transaction value.

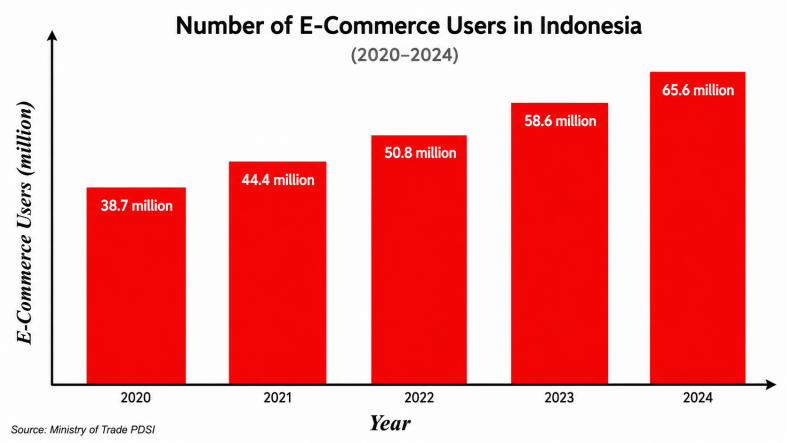


Figure 2. Number of e-commerce users in Indonesia.

Indonesian consumer behavior shows a high sensitivity to promotions. Survey findings indicate that promotions and discounts are dominant factors in determining consumers' choice of shopping platforms [2]. Empirical studies also show that digital coupons and discounts have a significant influence on purchase decisions and customers' perceived value [5], and can increase impulse buying behavior through hedonic motivation [6]. However, mass promotional practices without proper segmentation may lead to budget inefficiency and excessive promotional spending, which can reduce profit margins [7]. Consumer considerations when choosing an e-commerce platform are presented in Figure 3.

These issues highlight the importance of developing more data-driven and targeted promotional strategies. One approach that can be applied is customer segmentation, which helps identify differences in transaction behavior patterns [8–10]. Through segmentation, customers can be grouped based on characteristics such as transaction frequency, purchase value, recency, product variety, and voucher usage patterns. In this study, K-Means Clustering is used to identify customer groups with distinct transaction behaviors, allowing promotional strategies to be adjusted according to the characteristics of each segment [11, 12].

After the segmentation process, a predictive model is needed to estimate the probability of customers using targeted vouchers. This study employs two gradient boosting-based algorithms, namely XGBoost [13] and CatBoost [14], as these models perform well on tabular data and are capable of capturing non-linear relationships among variables [15, 16]. XGBoost has an advantage in model optimization through regularization, which helps reduce the risk

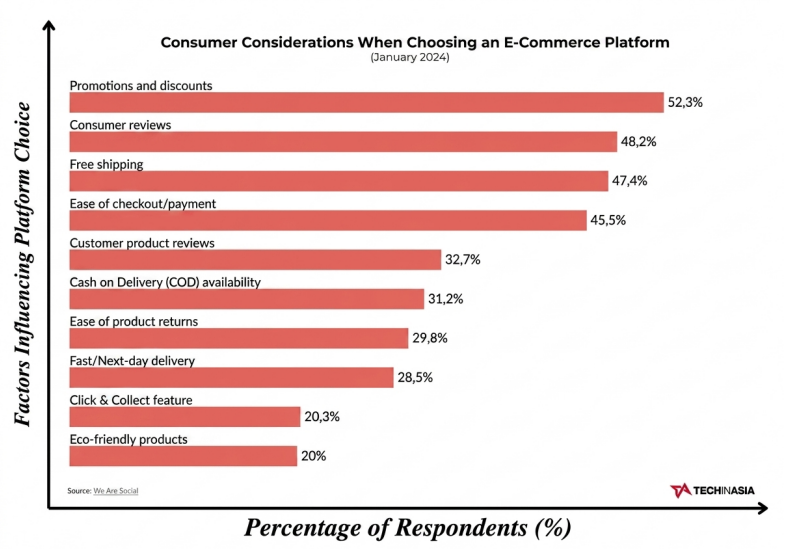


Figure 3. Consumer considerations when choosing an e-commerce platform.

of overfitting [17, 18], while CatBoost is particularly effective in handling categorical features directly. This is relevant because retail transaction data generally consist of a combination of numerical and categorical variables, such as transaction value, purchase quantity, payment method, product category, and customer characteristics [19, 20].

Although predictive models can estimate the likelihood of customers using vouchers, conventional classification approaches cannot explain whether voucher distribution actually causes customers to make a transaction. In other words, predictive models identify customers who are likely to redeem vouchers but cannot distinguish customers who would purchase regardless of promotional offers from those whose purchasing decisions are genuinely influenced by them. As a result, the incremental impact of voucher distribution remains unmeasured, highlighting the need for a causal inference approach.

To address this limitation, this study applies uplift modelling, a causal inference approach used to estimate the incremental effect of a treatment, such as a marketing intervention, on individual customer behavior [21]. Unlike conventional classification, uplift modelling compares predicted outcomes under treatment and control conditions to isolate the true causal effect of the intervention from customers' baseline purchasing behavior [22]. In this study, uplift modelling is implemented using a two-model approach, where separate predictive models are trained for the treatment and control groups, and the uplift score is calculated as the difference between their predicted probabilities [21].

Building upon this approach, this study integrates K-Means Clustering, XGBoost, CatBoost, and uplift modelling into a unified analytical framework. K-Means Clustering is employed to identify customer segments based on transaction behavior, while XGBoost and CatBoost are used to predict the probability of targeted voucher usage [13, 14]. These predictions are then incorporated into a two-model uplift framework to estimate the incremental effect of voucher distribution [21]. Consequently, promotional strategies can prioritize not only customers who are likely to redeem vouchers but also those whose purchasing behavior is genuinely influenced by voucher campaigns, resulting in more efficient targeting than mass campaigns or probability-based approaches.

Based on this background, this study aims to develop an integrated framework for customer segmentation, voucher usage prediction, and causal impact estimation. Specifically,

the study aims to segment customers using K-Means Clustering, predict targeted voucher usage using XGBoost and CatBoost, and estimate the incremental impact of voucher distribution through uplift modelling, thereby supporting more effective promotional targeting strategies.

The scope of this study is limited to growing-up milk product transactions from one e-commerce platform during the period from January to December 2024. The variables used in this study only include historical transaction data and customer characteristics available in the system, without incorporating external factors such as macroeconomic conditions or competitors' campaigns. This study focuses on two main approaches: unsupervised learning using K-Means Clustering for customer segmentation, and supervised learning using CatBoost and XGBoost combined with uplift modelling to predict customer responses and measure the causal impact of voucher distribution. Model performance evaluation is limited to ROC-AUC, PR-AUC, confusion matrix, and classification report, while promotional effectiveness is evaluated using uplift score, uplift ranking, and incremental gain.

Therefore, this study is expected to contribute to the development of more precise data-driven promotional strategies and serve as a reference for implementing data-driven marketing in the retail e-commerce sector.

2. Methods

2.1. Data and Preprocessing

The data used in this study consist of e-commerce transaction data from January to December 2024. The dataset includes both numerical and categorical variables, such as transaction value (sales), purchase quantity (qty), payment method, product category, customer demographics, and voucher usage information. The variables used in this study are presented in Table 1.

Table 1. Variables in the dataset.

Variable	Description	Measurement Scale
Y	Type of voucher used in the transaction	Nominal
X_1	Unique member identification number associated with the transaction	Nominal
X_2	Date on which the transaction was conducted	Interval
X_3	Unique transaction receipt number	Nominal
X_4	Product code of the purchased growing-up milk item	Nominal
X_5	Product name or description corresponding to the product code	Nominal
X_6	Category code of the purchased product	Nominal
X_7	Product category description corresponding to the category code	Nominal
X_8	Payment method used in the transaction	Nominal
X_9	Total sales value generated from the transaction	Ratio
X_{10}	Quantity of growing-up milk products purchased	Ratio
X_{11}	Store code indicating the branch where the transaction took place	Nominal
X_{12}	Special price discount applied to the transaction	Ratio
X_{13}	Final price after discount	Ratio
X_{14}	Gender of the registered member	Nominal
X_{15}	Date of birth of the registered member	Interval
X_{16}	Marital status of the registered member	Nominal
X_{17}	Application uninstall status of the member (Yes/No)	Nominal
X_{18}	Fraud indication status of the transaction (Yes/No)	Nominal
X_{19}	Return status of the transaction (Yes/No)	Nominal
X_{20}	Nominal value of the voucher applied in the transaction	Ratio

The preprocessing stage was conducted to ensure data quality before proceeding to the analysis stage. This process includes:

1. Data cleaning by removing returns or invalid transactions;
2. Handling missing values;

3. Data type conversion for date, numerical, and categorical variables; and
4. Standardization of numerical features using z-score normalization.

Standardization was applied to prevent large-scale features from dominating the distance calculation in K-Means [11]. The z-score standardization is given by

$$z = \frac{x - \mu}{\sigma}, \tag{1}$$

where x represents the original value, μ represents the mean, and σ represents the standard deviation of the feature.

2.2. Research Design

This study employs a quantitative machine learning-based approach to optimize voucher targeting strategies for customers of growing-up milk products. The research design is systematically structured by integrating three main analytical approaches: customer segmentation using unsupervised learning, predictive modelling using supervised learning, and causal modelling through uplift modelling.

The first stage aims to identify customer groups based on similarities in transaction behavior using K-Means Clustering [11]. The second stage focuses on developing classification models to predict the likelihood of targeted voucher usage using XGBoost [13] and CatBoost [14]. The third stage applies uplift modelling to measure the causal impact of voucher distribution on customer response [21]. The integration of these three stages enables this study not only to predict customer responses but also to measure promotional effectiveness incrementally. The overall research design is presented in Figure 4.

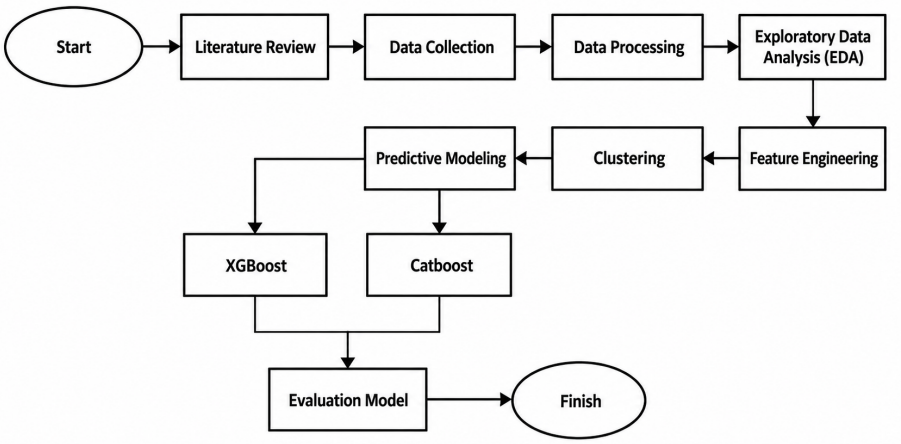


Figure 4. Research flowchart.

2.3. Feature Engineering

Feature engineering was conducted to create representative features that capture customer behavior patterns. This process involved aggregating transaction data at the customer level and constructing features such as:

- Recency, which represents the time since the last purchase;
- Frequency, which represents the number of transactions;
- Monetary value, which represents the total transaction value;
- Total quantity purchased;

- Average purchase value; and
- Voucher usage indicators.

This approach is aligned with the RFM concept in database marketing [23], which has been proven effective in measuring customer value. In addition, the cluster labels generated from the segmentation process were added as an additional feature (`cluster_kmeans`) in the predictive modelling stage to evaluate the contribution of segmentation to model performance.

2.4. Customer Segmentation with K-Means

Customer segmentation was performed using the K-Means Clustering algorithm to group customers based on similarities in transaction behavior [11, 12]. This approach aims to identify heterogeneity in customer behavior before conducting predictive modelling. K-Means works by minimizing the Within-Cluster Sum of Squares (WCSS):

$$J = \sum_{j=1}^{K} \sum_{x_i \in C_j} \|x_i - \mu_j\|^2, \tag{2}$$

where J is the Within-Cluster Sum of Squares (WCSS), K is the total number of clusters, C_j denotes the set of data points assigned to cluster j, x_i is the i-th data point, and μ_j is the centroid of cluster j.

The optimal number of clusters was determined using the Silhouette Score:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \tag{3}$$

where $a(i)$ represents the average distance between data point i and other points within the same cluster, while $b(i)$ represents the minimum average distance between data point i and points in the nearest neighboring cluster. A Silhouette Score close to 1 indicates that the clusters are well separated. The clustering results were then used as an additional feature (`cluster_kmeans`) in the predictive modelling stage to evaluate the contribution of segmentation in improving model performance.

2.5. Predictive Modelling

The predictive modelling stage was formulated as a binary classification problem [24] to predict whether a customer would redeem a targeted voucher or not. The target variable is defined as

$$y = \begin{cases} 1, & \text{if the transaction uses a targeted voucher,} \\ 0, & \text{otherwise.} \end{cases}$$

The dataset was divided into training and testing sets using a stratified train-test split to maintain the class proportion. Two gradient boosting-based algorithms were used, namely XGBoost [13] and CatBoost [14].

2.5.1. XGBoost

XGBoost optimizes an objective function consisting of a loss function and a regularization term [13]:

$$L = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i\right) + \sum_{k=1}^{K} \Omega(f_k), \tag{4}$$

where L is the overall objective function of XGBoost, n is the total number of training samples, $l(y_i, \hat{y}_i)$ is the loss function that measures the difference between the true label y_i and the predicted value $\hat{y}_i$, y_i is the true label of the i-th training sample, $\hat{y}_i$ is the predicted value for the i-th training sample, K is the total number of trees in the ensemble, f_k is the prediction function of the k-th decision tree, and $\Omega(f_k)$ is the regularization term that penalizes the complexity of the k-th tree.

Regularization helps control model complexity and prevent overfitting [13].

2.5.2. CatBoost

CatBoost is designed to handle categorical features directly using ordered boosting [14], so it does not require explicit one-hot encoding and helps reduce the risk of target leakage. This is relevant because the transaction dataset contains several categorical features, such as payment method and product category.

2.6. Model Evaluation

Model classification performance evaluation was conducted to measure the extent to which the model was able to distinguish transactions using targeted vouchers from non-targeted transactions. Since the dataset had an imbalanced class distribution, where non-targeted transactions were significantly more dominant, the selection of appropriate evaluation metrics was crucial. Therefore, this study employed ROC–AUC, PR–AUC, Confusion Matrix, and Classification Report as the main evaluation tools [25].

2.6.1. Receiver Operating Characteristic–Area Under Curve (ROC–AUC)

ROC–AUC was used to evaluate the model's overall ability to distinguish between the positive and negative classes without relying on a specific threshold value [25]. The ROC curve represents the relationship between:

- True Positive Rate (TPR), or Recall; and
- False Positive Rate (FPR).

TPR and FPR are formulated as follows:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{5}$$

and

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}. \tag{6}$$

2.6.2. Precision–Recall Area Under Curve (PR–AUC)

In imbalanced datasets, ROC–AUC may sometimes be less sensitive to the model's performance on the minority class. Therefore, the Precision–Recall AUC (PR–AUC) metric was used, as it focuses more specifically on the performance of the positive class. Precision is defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{7}$$

Precision measures how many positive predictions are actually relevant. In a business context, a high precision value indicates that most customers predicted as voucher candidates are truly likely to use the voucher, thereby reducing unnecessary promotional budget allocation.

Recall is formulated as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{8}$$

Recall measures the model's ability to identify all customers who actually used the voucher. A high recall value indicates that only a small number of potential customers are missed by the model.

2.6.3. Confusion Matrix

The Confusion Matrix provides a detailed representation of the model's classification results in the form of a 2×2 matrix consisting of TP, TN, FP, and FN. This matrix enables a direct analysis of the types of errors produced by the model.

Accuracy is formulated as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{9}$$

Accuracy measures the overall proportion of correct predictions. However, in imbalanced datasets, accuracy may be biased because the model can achieve a high score by dominantly predicting the majority class.

The F1-Score is the harmonic mean of precision and recall:

$$\text{F1} = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{10}$$

F1-Score was used to evaluate the balance between precision and recall at a specific threshold. In this study, a threshold value of 0.5 was applied. The Classification Report was then used to comprehensively compare the performance of XGBoost and CatBoost before proceeding to the uplift modelling stage.

2.7. Uplift Modelling

Uplift modelling was used to measure the causal impact of targeted voucher allocation on changes in customers' redemption probability [21]. Unlike conventional classification, which only predicts the likelihood of a purchase or redemption, uplift modelling estimates the incremental effect between customers who receive the promotion as the treatment group and those who do not receive the promotion as the control group.

In this study, a quasi-experimental approach was applied using historical transaction data. The treatment variable, `T_eligible`, was determined based on the results of predictive modelling, where customers in the top 30% with the highest predicted probabilities were classified as eligible customers ($T = 1$), while the remaining customers were classified as non-eligible customers ($T = 0$). Meanwhile, the response variable, `Y_post`, was defined as an indicator of targeted voucher redemption during the evaluation period.

The method used in this study was the two-model approach [21], which involves developing two separate models:

$$P(Y = 1 \mid T = 1, x), \tag{11}$$

and

$$P(Y = 1 \mid T = 0, x). \tag{12}$$

The uplift value is calculated as follows:

$$U(x) = P(Y = 1 \mid T = 1, x) - P(Y = 1 \mid T = 0, x). \tag{13}$$

The higher the value of $U(x)$, the greater the incremental effect of providing a voucher to that customer.

Uplift evaluation was conducted using the following approaches:

- Uplift ranking, to determine customer targeting priorities based on the estimated incremental effect;
- Decile analysis, to measure how the incremental impact is concentrated across customer segments;
- Incremental effect analysis, calculated as the difference in redemption rates between the treatment and control groups within the highest uplift segment; and
- Strategy comparison, including mass campaign, predictive targeting using the top 30% probability group, and uplift-based targeting.

This approach enables the identification of customers categorized as persuadables, namely customers whose probability of voucher redemption truly increases as a result of receiving the voucher [21, 22]. Therefore, the targeting strategy can be implemented more efficiently compared to an approach that relies solely on predicted probability.

3. Results and Discussion

This section presents the research findings obtained from all stages of analysis, including data understanding and cleaning, exploration of customer characteristics, customer segmentation using K-Means Clustering, evaluation of the XGBoost and CatBoost predictive models, and uplift modelling analysis. The discussion in this section focuses not only on the technical performance of the models but also on the interpretation of the results within the context of e-commerce retail promotion strategies. Therefore, the findings can be used to assess whether an approach based on segmentation, predictive modelling, and uplift modelling is able to produce a more targeted voucher strategy compared to mass promotion.

3.1. Data Overview and Preprocessing

The research dataset consisted of 2,204,504 retail transactions for growing-up milk products, with 21 variables covering transaction information, customer characteristics, and voucher usage. After the preprocessing stage, which included handling missing values, filtering returned transactions, cleaning anomalous values, and restricting the customer age range, the final dataset contained 2,166,552 valid transactions for further analysis.

Before conducting the main analysis, data preprocessing was performed to ensure the quality and reliability of the dataset. This stage included missing value handling, data type transformation, derived variable construction, and the removal of invalid transactions. A summary of the preprocessing steps is presented in Table 2.

3.2. Exploratory Data Analysis

The exploratory data analysis (EDA) stage was conducted to understand the distributional characteristics of the data, customer behavior patterns, and potential class imbalance prior to model development. This analysis covered the demographic distribution of customers, transaction patterns across generations, and the distribution of voucher types used in the transactions.

Based on the demographic analysis, the majority of transactions were made by female customers, accounting for 79.81% of total transactions, while male customers accounted for only 20.19%. This finding indicates that growing-up milk products were predominantly purchased by female customers, which may be associated with their role in making purchasing

Table 2. Summary of data preprocessing stages.

Preprocessing Step	Applied Treatment	Objective
Missing value handling	Missing values in the `fraud` variable were filled with `N`, while missing values in `nominal_voucher` were filled with 0.	To maintain data consistency and represent non-fraud conditions as well as the absence of voucher usage.
Data type conversion	The `tgl` and `tanggal_lahir` variables were converted into datetime format.	To enable time-based analysis and customer age calculation.
Age calculation	Customer age was calculated based on the difference between `tanggal_lahir` and the end of the observation period.	To construct a demographic feature representing customer age.
Age imputation	Missing age values were imputed using the median value.	To reduce data loss caused by missing values.
Construction of `fix_price`	The actual unit price was calculated using `harga`, `dh_spesial`, and `qty`.	To obtain a net price representation after considering discounts and purchase quantity.
Standardization of `tipe_voucher`	Missing values in `tipe_voucher` were filled with `no_voucher` and converted into string format.	To maintain consistency across voucher categories.
Targeted voucher encoding	A binary variable, `is_targeted`, was created based on the use of targeted vouchers.	To serve as the target variable in the predictive modelling stage.
Voucher usage encoding	A binary variable, `is_any_voucher`, was created to indicate whether a transaction used any type of voucher.	To support the analysis of voucher usage behavior.
Age group classification	Customer age was grouped into the `generation` variable.	To support demographic analysis of customers.
Discount indicator construction	The `discount_flag` variable was created based on the value of `dh_spesial`.	To identify transactions involving promotions or discounts.
Transaction filtering	Returned transactions and invalid values in `qty`, `sales`, and `fix_price` were removed.	To ensure that only valid transactions were included in the analysis.
Age range restriction	Customer data were restricted to the age range of 15–70 years.	To reduce the influence of anomalous age values on the analysis.

decisions for children's needs. The gender distribution of customers is presented in Figure 5.

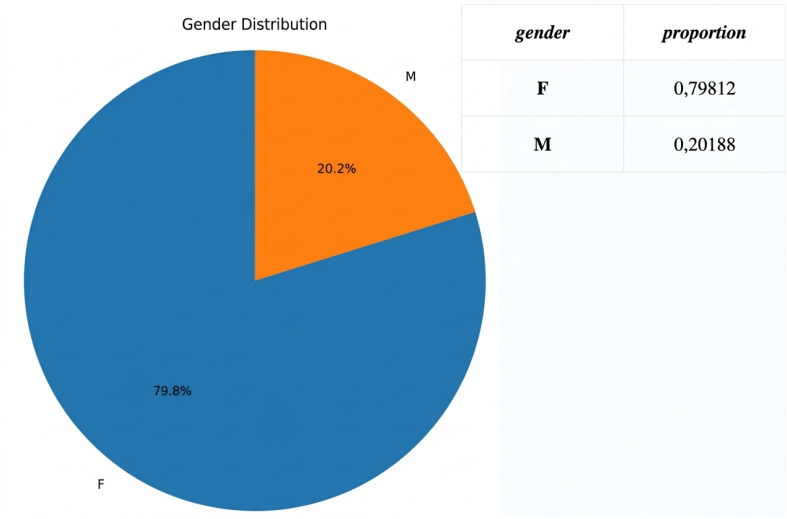
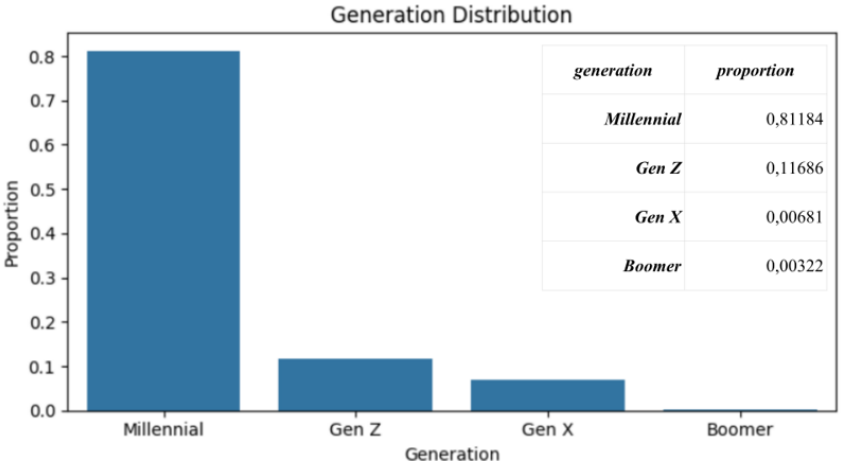


Figure 5. Proportion of member gender.

Furthermore, the transaction distribution by generation shows that 81.18% of customers belonged to the Millennial generation, followed by other generations with substantially smaller proportions. This indicates that customers within the productive age group were the main contributors to growing-up milk sales on the analysed e-commerce platform. The dominance of Millennials also suggests a high level of digital adoption among this group in online shopping activities. The distribution of customers by generation is shown in Figure 6.

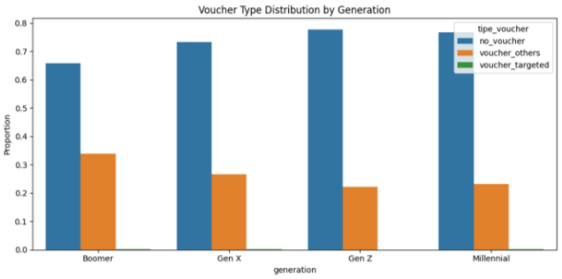


generation	*proportion*
Millennial	0,81184
Gen Z	0,11686
Gen X	0,00681
Boomer	0,00322

Figure 6. Proportion of member generation.

From the perspective of voucher usage, the analysis shows that targeted voucher redemption was a rare event, accounting for less than 0.25% of total transactions. Most transactions were completed without using any voucher, with a proportion exceeding 70%. This characteristic confirms that the classification task in this study falls under an imbalanced classification problem, where the positive class, namely targeted voucher usage, is substantially smaller than the negative class.

The distribution of voucher types across generations shows a relatively consistent pattern, in which most customers did not use vouchers regardless of their age group. This indicates that targeted voucher usage is not solely influenced by a single demographic factor, such as age or gender, but is more likely affected by a combination of historical transaction behavior, purchase frequency, and spending value. The distribution of voucher types across generations is presented in Figure 7.



generation	*tipe_voucher*	*count*	*proporsi*
Boomer	*no_voucher*	4677	0,65836
Boomer	*voucher_others*	2411	0,33939
Boomer	*voucher_targeted*	16	0,00225
Gen X	*no_voucher*	109941	0,73255
Gen X	*voucher_others*	39952	0,26621
Gen X	*voucher_targeted*	187	0,00125
Gen Z	*no_voucher*	200378	0,77782
Gen Z	*voucher_others*	57152	0,22185
Gen Z	*voucher_targeted*	85	0,00033
Millennial	*no_voucher*	1373521	0,76746
Millennial	*voucher_others*	414176	0,23142
Millennial	*voucher_targeted*	2008	0,00112

Figure 7. Proportion of voucher types by generation.

3.3. Customer Segmentation (Clustering)

Customer segmentation was conducted using the K-Means Clustering algorithm to identify groups of customers with similar historical transaction behavior. This approach aimed to capture customer heterogeneity prior to predictive modelling, allowing the model to incorporate more granular behavioral structures.

The optimal number of clusters was determined by evaluating the Silhouette Score for values of K ranging from 2 to 8. The evaluation results showed that the highest silhouette

score was achieved at $K = 7$, with a score of 0.246; therefore, seven clusters were selected as the optimal cluster configuration. Although the silhouette score was not close to 1, this is commonly observed in large-scale transactional behavior data with high variability. Nevertheless, the score still indicates that the resulting clusters provide a sufficiently representative separation in the context of retail customer segmentation. The Silhouette Score evaluation is presented in Figure 8.

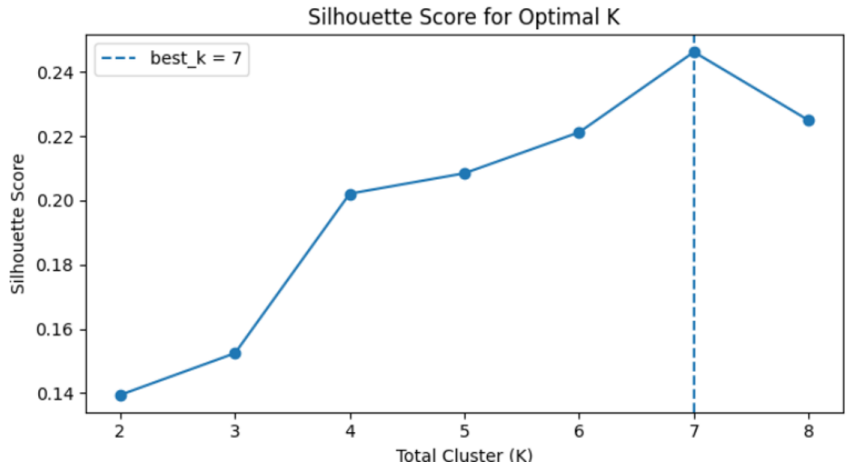


Figure 8. Silhouette score for K-Means.

The visualization of the customer clustering results in a two-dimensional projection is presented in Figure 9.

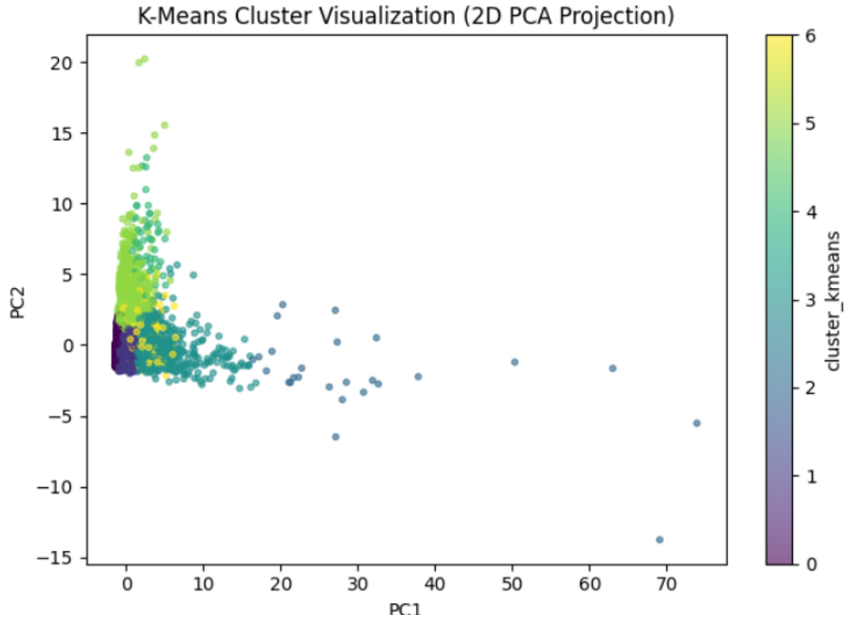


Figure 9. Visualization of customer clustering results in a two-dimensional projection.

The clustering results produced seven customer groups with different behavioral characteristics. In general, the cluster structure can be interpreted as presented in Table 3.

This segmentation reveals clear differences between high-value customers and passive customers, both in terms of transaction intensity and promotional responsiveness. Such heterogeneity indicates that a mass campaign-based promotional strategy may not be optimal, as customer needs and sensitivity toward vouchers differ across clusters.

Table 3. K-Means cluster description.

Cluster	Main Characteristics	Behavioral Interpretation
Cluster 0	Very low transaction frequency, low total sales, high recency, and almost no voucher usage	Low-activity customers, referring to passive customers who rarely make transactions and show limited responsiveness to promotions.
Cluster 1	Moderate transaction frequency, medium sales value, and low recency	Moderate transaction frequency, medium sales value, and low recency.
Cluster 2	Very high transaction frequency, extremely high total sales, and highly diverse SKU and branch coverage	Very high transaction frequency, extremely high total sales, and highly diverse SKU and branch coverage.
Cluster 3	Relatively high transaction frequency, moderate product variety, and low recency	Relatively high transaction frequency, moderate product variety, and low recency.
Cluster 4	High average sales value per transaction and evidence of voucher usage	High average sales value per transaction and evidence of voucher usage.
Cluster 5	Low to medium transaction frequency and relatively high recency	Low to medium transaction frequency and relatively high recency.
Cluster 6	Medium transaction frequency and active purchasing behavior across multiple branches	Medium transaction frequency and active purchasing behavior across multiple branches.

3.4. Predictive Modelling

The predictive modelling stage was formulated as a binary classification problem to predict the probability of customers redeeming targeted vouchers. Two gradient boosting-based algorithms, namely XGBoost and CatBoost, were employed in this study due to their ability to handle large-scale tabular data and capture non-linear relationships among features.

To prevent data leakage, the dataset was divided using a time-based split with a cutoff date of 1 November 2024. Thus, the models were trained using historical data and tested on a subsequent evaluation period. Model performance was evaluated using ROC–AUC and PR–AUC as measures of discriminative ability, along with threshold-based metrics such as precision, recall, F1-score, and accuracy at a cutoff value of 0.5.

Considering that targeted voucher redemption is a rare event, accounting for less than 0.25% of total transactions, PR–AUC and F1-score are considered more relevant indicators than accuracy in evaluating model performance. The evaluation results are presented in Table 4.

Table 4. Evaluation of ROC–AUC, PR–AUC, and classification report.

Model	Cluster	ROC–AUC	PR–AUC	Precision	Recall	F1-score	Accuracy
XGBoost	No	0.9960	0.4236	0.4980	0.1123	0.1833	0.9974
CatBoost	No	0.9960	0.3793	0.4255	0.0731	0.1247	0.9973
XGBoost	Yes	0.9972	0.5479	0.6306	0.1918	0.2941	0.9976
CatBoost	Yes	0.9966	0.5262	0.5586	0.1653	0.2551	0.9975

Based on Table 4, the addition of the `cluster_kmeans` feature improved the performance of both models, particularly in terms of PR–AUC and F1-score. This indicates that customer segmentation is not only useful for understanding customer characteristics but also provides additional relevant information for predictive modelling. By incorporating the cluster feature, the model can consider broader customer behavioral patterns, such as transaction intensity, spending value, recency, and tendency to use vouchers.

The improvement in PR–AUC is particularly important because targeted voucher redemption represents the minority class in the dataset. In imbalanced data conditions, accuracy tends to be less representative, as a model may appear to perform well simply by predicting most transactions as non-targeted. Therefore, the increase in PR–AUC and F1-score indicates that the model is better able to identify transactions or customers that are truly relevant for targeted voucher allocation.

Overall, XGBoost with the cluster feature achieved the best performance compared to the other configurations. This result suggests that combining customer segmentation with a boosting-based predictive model can produce a more precise targeting approach than models that rely solely on transactional features without segmentation information.

The confusion matrix evaluation for each model configuration is presented in Table 5.

Table 5. Confusion matrix evaluation.

Model	Cluster	TN	FP	FN	TP
XGBoost	No	422,383	124	972	123
CatBoost	No	422,399	108	1,015	80
XGBoost	Yes	422,384	123	885	210
CatBoost	Yes	422,364	143	914	181

Table 5 presents the confusion matrix of each classification model, which provides a detailed view of the prediction results. True Negative (TN) represents observations that were correctly predicted not to redeem the targeted voucher, False Positive (FP) represents observations that were incorrectly predicted to redeem the targeted voucher, False Negative (FN) represents observations that actually redeemed the targeted voucher but were predicted not to redeem it, and True Positive (TP) represents observations that were correctly predicted to redeem the targeted voucher.

The results show that all models achieved a very high number of true negatives, with more than 422,000 observations, indicating that they were able to accurately identify the majority class. This is expected because the dataset is highly imbalanced, where targeted voucher redemption represents only a small proportion of all observations. Consequently, the differences between models are more evident in their ability to identify the minority (positive) class.

The addition of the cluster feature improved the models' capability to detect targeted voucher redemption. For XGBoost, the number of true positives increased from 123 to 210, while false negatives decreased from 972 to 885 after incorporating cluster information. Similarly, CatBoost increased its true positives from 80 to 181 and reduced false negatives from 1,015 to 914. These improvements indicate that customer segmentation provides additional behavioral information that enables both models to better distinguish observations associated with targeted voucher redemption.

However, this improvement comes with a slight trade-off. The number of false positives increased marginally for CatBoost, from 108 to 143, while XGBoost remained almost unchanged, from 124 to 123. In practical terms, false positives represent customers who are predicted as likely to redeem vouchers despite not actually redeeming them, potentially increasing promotional costs if they are selected for targeting. On the other hand, false negatives represent missed opportunities because customers who actually redeem vouchers are not identified by the model. Therefore, reducing false negatives is generally more desirable, provided that false positives do not increase excessively.

Overall, XGBoost with the cluster feature provides the most balanced performance. Compared with the non-cluster model, it substantially increases the number of correctly identified targeted voucher redemptions while keeping false positives at nearly the same level. This result is consistent with the higher F1-score and PR–AUC reported in Table 4, indicating that customer clustering contributes meaningful information for improving targeted voucher prediction.

Although the classification results demonstrate strong predictive performance, this ap-

proach is still limited to estimating the probability of voucher usage without measuring the causal impact of providing the voucher. The model cannot distinguish between customers who would have purchased without a promotion and customers whose behavior was actually influenced by the promotion. Therefore, an uplift modelling approach is required to estimate the incremental effect of voucher allocation, which is discussed in the following section.

3.5. Uplift Modelling

This study further extends the analysis by applying an uplift modelling approach to estimate the incremental effect of voucher allocation on customers' redemption behavior.

The approach used in this study was the two-model approach, in which separate models were developed to estimate the probability of redemption under the treatment and control conditions. The treatment variable, `T_eligible`, was determined based on the predictive modelling results, where customers in the top 30% of predicted probabilities were categorized as eligible customers ($T = 1$), while the remaining customers were categorized as non-eligible customers ($T = 0$). The response variable, `Y_post`, was defined as an indicator of targeted voucher redemption during the evaluation period.

The distribution of the treatment and response groups is presented in Table 6.

Table 6. Distribution of treatment and response.

Category	Number of Customers	Redemption ($Y = 1$)	No Redemption ($Y = 0$)	Redemption Rate
Treatment (Top 30%)	66,611	1,033	53,450	1.91%
Control (Bottom 70%)	155,425	6	127,121	0.00%
Total	222,036	1,039	180,571	0.47%

Table 6 shows that the reported redemption rate in the treatment group was 1.91%, substantially higher than that of the control group, which was close to zero. Under the model-defined treatment and control framework, this difference corresponds to an estimated uplift of 1.91%. In aggregate, the uplift analysis identified approximately 1,027 incremental redemptions associated with the selected customer group.

To further understand the distribution of uplift effects conceptually, customers were then mapped into four main uplift modelling categories, as presented in Table 7.

Table 7. Distribution of the four customer types.

Customer Type	Estimated Number	Proportion	Characteristics	Strategic Implication
Persuadables	$\approx 1{,}027$	0.46%	Customers who redeem or purchase because they receive a voucher	Customers who redeem or purchase because they receive a voucher
Sure Things	≈ 6	$\approx 0.00\%$	Customers who would redeem or purchase even without a voucher	Customers who would redeem or purchase even without a voucher
Lost Causes	$\approx 220{,}997$	99.54%	Customers who do not redeem or purchase even after receiving a voucher	Customers who do not redeem or purchase even after receiving a voucher
Sleeping Dogs	Very small / not significantly detected	$\approx 0\%$	Customers who respond negatively to voucher intervention	Customers who respond negatively to voucher intervention

4. Conclusion

This study aimed to optimize voucher targeting strategies for retail transactions of growing-up milk products by integrating customer segmentation, predictive modelling, and uplift modelling. The customer segmentation process using K-Means produced seven clusters with het-

erogeneous behavioral characteristics, which were shown to improve classification performance when incorporated as an additional feature. In the predictive modelling stage, XGBoost with the cluster feature achieved the best performance, with a ROC–AUC of 0.9972 and a PR–AUC of 0.5479, along with improved precision and recall compared to models without segmentation.

Although the classification model was able to accurately predict redemption probability, this approach did not measure the incremental effect of voucher allocation. Therefore, uplift modelling was applied and identified approximately 1,027 customers, or 0.46% of the total 222,036 customers, as persuadables under the model-based uplift framework. This segment generated a reported uplift value of 1.91%.

These findings indicate that an uplift-based targeting strategy has the potential to be more efficient than both mass campaigns and probability-based predictive targeting, as it enables promotional efforts to be focused on customers with the highest estimated incremental impact. Thus, the integration of customer segmentation, predictive modelling, and uplift modelling provides a more precise and uplift-informed approach to optimizing e-commerce promotional strategies.

Author Contributions. **Michael Calvin Charmelino Wijaya**: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing—original draft preparation, writing—review and editing, and visualization. **Maria Zefanya Sampe**: Conceptualization, methodology, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, and supervision.

Acknowledgment. The author would like to express sincere gratitude to all parties who contributed to this research and to the preparation of this article. The author also greatly appreciates the editors and reviewers for their valuable feedback and support in improving this work.

Funding. This research received no external funding.

Data availability. The data are not publicly available due to privacy or institutional restrictions.

Conflict of Interest. The authors declare that there are no conflicts of interest related to this article.

References

[1] International Trade Administration, "E-commerce sales size & forecast," Trade.gov, 2024, https://www.trade.gov/ecommerce-sales-size-forecast.
[2] Tech in Asia, "Data e-commerce indonesia: Panduan lengkap," Tech in Asia, 2025.
[3] ECDB, "Indonesian ecommerce market revenues projected to cross us$100 billion in 2025," 2023.
[4] Google, Temasek, and Bain & Company, "e-conomy sea 2024: The rise of ai and digital economy," 2024.
[5] D. Kavitha, T. Varalaxmi, and E. Mamatha, "An empirical study on customers' preferences towards digital marketing strategy: Coupon and discount based promotional activities," *Asian Journal of Multidimensional Research*, vol. 9, no. 1, pp. 3730–3735, 2020.
[6] J. Fathoni, "The effect of discounts, cashback and promotions on impulse buying through hedonic shopping motivation as an intervening variable in shopee consumers islamic perspective: A case study of students in banyuwangi regency," *Indonesian Journal of Islamic Economics and Finance*, vol. 7, no. 1, Jun 2024, doi:10.35719/ijief.v7i1.1545.
[7] Boston Consulting Group, "How to run effective retail promotions in saturated markets," 2023.
[8] P. Kotler, K. L. Keller, and A. Chernev, *Marketing Management*, 17th ed. Pearson, 2025, published by Pearson on 6 November 2024.

[9] P. Kotler, G. Armstrong, and S. Balasubramanian, *Principles of Marketing*, 19th ed. Pearson, 2024, published by Pearson on 20 July 2023.
[10] M. A. Gomes and T. Meisen, "A review on customer segmentation methods for personalized customer targeting in e-commerce use cases," *Information Systems and e-Business Management*, vol. 21, no. 3, pp. 527–570, 2023, doi:10.1007/s10257-023-00640-4.
[11] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967.
[12] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 1990, doi:10.1002/9780470316801.
[13] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug 2016, pp. 785–794, doi:10.1145/2939672.2939785.
[14] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in *Advances in Neural Information Processing Systems*, 2018, https://proceedings.neurips.cc/paper_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf.
[15] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on tabular data?" *arXiv preprint arXiv:2207.08815*, 2022, doi:10.48550/arXiv.2207.08815.
[16] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *arXiv preprint arXiv:2106.03253*, 2021, doi:10.48550/arXiv.2106.03253.
[17] W. Wang, W. Xiong, J. Wang, L. Tao, S. Li, Y. Yi, X. Zou, and C. Li, "A user purchase behavior prediction method based on XGBoost," *Electronics*, vol. 12, no. 9, p. 2047, 2023, doi:10.3390/electronics12092047.
[18] S. C. Sapkota, P. Saha, S. Das, and L. V. P. Meesaraganda, "Prediction of the compressive strength of normal concrete using ensemble machine learning approach," *Asian Journal of Civil Engineering*, vol. 25, no. 1, pp. 583–596, 2024, doi:10.1007/s42107-023-00796-x.
[19] L. Gan, "XGBoost-based e-commerce customer loss prediction," *Computational Intelligence and Neuroscience*, vol. 2022, p. 1858300, 2022, doi:10.1155/2022/1858300.
[20] P. K. Ram, N. Pandey, and J. Persis, "Modeling social coupon redemption decisions of consumers in food industry: A machine learning perspective," *Technological Forecasting and Social Change*, vol. 200, p. 123093, 2024, doi:10.1016/j.techfore.2023.123093.
[21] H. Langen and M. Huber, "How causal machine learning can leverage marketing strategies: Assessing and improving the performance of a coupon campaign," *arXiv preprint arXiv:2204.10820*, 2022, doi:10.48550/arXiv.2204.10820.
[22] Y. T. Park, T. Xu, and M. Anany, "Enhancing uplift modeling in multi-treatment marketing campaigns," *arXiv preprint arXiv:2408.13628*, 2024, doi: https://arxiv.org/pdf/2408.13628.
[23] A. M. Hughes, *Strategic Database Marketing: The Masterplan for Starting and Managing a Profitable, Customer-Based Marketing Program*, 4th ed. McGraw-Hill, 2012, https://www.scribd.com/document/190459529/32792977-Summary-Strategic-Database-Marketing-Arthur-M-Hughes-pdf.
[24] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006, doi:10.1007/978-0-387-45528-0.
[25] T. M. Mitchell, *Machine Learning*. McGraw-Hill, 1997, http://www.cs.cmu.edu/~tom/files/MachineLearningTomMitchell.pdf.