Journal of Informatics and Web Engineering
Vol. 5 No. 3 (October 2026) eISSN: 2821-370X

A Comparative Study of Machine Learning Algorithms for Customer Churn Prediction in e-Commerce

Tan Yi Qi[1], Lew Sook Ling[2*], Ram Avtar[3]**
[1]Faculty of Information Science and Technology, Multimedia University, 75450 Melaka, Malaysia
[2]Centre for Advanced Analytics, CoE for Artificial Intelligence, FIST, MMU, 75450 Melaka, Malaysia
[3]Faculty of Environmental Earth Science, Hokkaido University, Sapporo 060-0810, Japan
**corresponding author: (sllew@mmu.edu.my; ORCiD: 0000-0003-4515-1163)*
***corresponding author: (ram@ees.hokudai.ac.jp; ORCiD: 0000-0003-3653-5771)*

Abstract - Customer churn rate is one of the main challenges that e-Commerce businesses must address, as it may affect revenue and long-term growth. In this paper, predictive analysis was used to identify potential customers who would be lost. It uses two different types of data to evaluate and compare the performance of five Machine Learning (ML) models, namely logistic regression, decision tree, support vector machine, gradient boost and random forest. Accuracy, precision, recall rate and F1 score are used as indicators of model performance. Among them, random forest achieved the highest performance across both datasets, demonstrating superior predictive power and consistency. By validating results on multiple datasets, this paper not only ensures reliability but also highlights the stability of churn prediction models across varying data conditions. The findings indicate that effective churn prediction can support proactive customer retention strategies, enabling e-Commerce businesses to minimize losses and promote sustainable growth. The use of two datasets represents a notable contribution, as it provides a stronger basis for assessing model generalizability and avoiding bias toward a single data source. The findings indicate that accurate churn prediction can support proactive retention strategies, enabling e-Commerce businesses to reduce churned customers, enhance loyalty, and maintain a competitive advantage. Overall, this paper underscores the value of data-driven approaches, demonstrates the benefits of employing multiple datasets for model validation, and provides actionable insights for businesses aiming to implement effective and sustainable customer retention strategies.

Keywords— Customer Churn, Customer Retention, e-Commerce, *Machine Learning, Predictive Analysis.*

Received: 16 July 2025; Accepted: 7 January 2026; Published: 16 October 2026

This is an open access article under the CC BY-NC-ND 4.0 license.

1. INTRODUCTION

E-Commerce has transformed the global retail landscape by enabling businesses to reach wider markets and provide personalized services. With the rise of digitalization and increasing competition, customer retention has become a critical challenge. One of the most pressing issues is customer churn, as acquiring new customers is often more costly

than retaining existing ones [1]. Predicting churn allows companies to proactively address dissatisfaction and design targeted retention strategies, ultimately improving business performance.

Although churn prediction has been extensively studied in sectors such as telecommunications, banking, and retail, e-Commerce presents unique challenges due to rapidly changing products, diverse customer preferences, and intense competition [1]. These complexities make Machine Learning (ML) techniques particularly valuable, as they can uncover hidden patterns in customer behaviour and identify those at risk of leaving. Prior research has shown that ensemble models such as Random Forest (RF) and Gradient Boosting achieve strong predictive accuracy in churn prediction tasks, while Deep Learning (DL) approaches like Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) provide temporal and spatial insights into customer behaviour [2]. However, many existing studies rely on a single dataset or focus on narrow industry contexts, limiting the generalizability of their findings [3]. At the same time, e-Commerce has become highly convenient, particularly for customers with limited time, as online platforms enable fast transactions and global business expansion [4]. With the rise of "Internet +" e-Commerce, traditional enterprises are increasingly relying on network platforms to sustain growth, and internet companies have fuelled the rapid development of the online retail market [4]. Despite these advantages, e-Commerce businesses face significant challenges related to customer churn. The vast volume of data makes it difficult for many companies to identify which customers are leaving, underscoring the importance of applying data mining and ML methods to predict churn more effectively.

This paper addresses this gap by applying and comparing five ML models across on two e-Commerce datasets. By doing so, the paper evaluates model stability across different datasets and provides insights into factors influencing customer churn in e-Commerce. The objective is (i) To develop a predictive model to identify customer churn risk based on customer buying behaviour and engagement data. (ii) To analyse the key factors to the loss of e-Commerce customers and develop targeted customer retention strategies. (iii) To evaluate the effectiveness of predictive models in reducing attrition rates and improving customer retention outcomes for e-Commerce businesses.

2. LITERATURE REVIEW

2.1 Churn Prediction

Churn prediction is to use ML to predict potential customers who may leave. Its main function is to quickly identify customers' dissatisfaction or restrict their service consumption behaviour. This is how businesses retain customers. The previous paper shows that the prediction of customer churn can help businesses to increase their revenue and market share [5]. Marketing to reach new customers is the most expensive and reaching new customers is overrated. Instead, it is cheaper to retain existing customers and keep customers churn to the minimum [5]. In fact, this practice can also increase shareholder value.

Customer churn prediction is a key link in businesses customer relationship management. It determines the possibility of customer trends based on customers' behavioural patterns and product usage, thereby issuing warnings to businesses. Through this technology, retail companies can also better determine when customers may leave, thereby taking effective actions in a timely manner to minimize customer churn rates and maximize customer retention rates.

ML technology can be used to collect customer data and analyse it to identify customer behaviours and habits, such as product usage behaviour, demographic segmentation and transaction history, and apply them to predict customer churn. This model will be able to identify the patterns and trends of customer churn. These models can not only predict the possibility of individual customer churn but also that of the entire customer base, thereby providing data to help businesses formulate customer maintenance strategies for specific units.

2.2 Machine Learning

Artificial Intelligence (AI) involves a technique known as ML to perform activities. It is mostly utilized to find some patterns or forecast the future scenario out of the statistics. The combination of AI and ML allows measuring the data more correctly, finding patterns, and reshaping predictions results, which will lead to more adequate services [6]. ML models can also help companies track potential customers they may lose and take immediate measures to retain them [7].

Through ML and DL technologies, e-Commerce businesses can proactively identify and address customer churn issues [8]. This can increase revenue and customer retention rates [8]. ML, neural networks and DL are just a few subfields under AI. Neural networks are a part of ML, and DL is a subset of neural networks [9].

Compared with other methods, text and images are examples of raw and unstructured data that DL can directly handle. It can analyse a large amount of data by itself and discover patterns in the data with almost no human help. However, traditional ML requires manual selection of important features and well-structured data. On the other hand, DL is more intelligent and effective and can identify beneficial features by itself. In this paper, the use of ML in predicting customer churn in e-Commerce to allow an earlier identification of risky customers to facilitate data driven retention strategies. Recent studies also show that handling class imbalance through oversampling techniques like SMOTE can further improve the predictive performance of ML and DL models by generating synthetic churn instances to balance the dataset [10].

2.3 Machine Learning Algorithm

ML algorithms are a set of rules or processes used by AI systems to perform tasks. Its most common task is to discover patterns in the data or predict future outcomes. When AI and ML are combined, they can help analyse data more precisely, identify patterns, and adjust predictions to provide more accurate services [11]. Therefore, ML can predict which customers are likely to churn by analysing past customer behaviour. In this paper, will apply ML algorithms to develop a predictive model which would help in predicting possible churn and allow the enterprises to know when to take any retention measures.

2.3.1 Random Forest (RF)

This paper focuses on predicting customer churn in e-Commerce. It can facilitate the company to know the number of customers who are still using the service at any time. The paper employed six ML algorithms namely RF, Linear Discriminant Analysis (LDA), Logistic Regression *(*LR), Decision Tree (DT), K-Nearest Neighbours (KNN), and Extreme Gradient Boosting (XGBoost). The result indicates that the highest accuracy level is achieved by the RF, reaching 98%. On the other hand, LR and LDA have an accuracy rate of 85% and 96%, KNN has the accuracy rate of 91%, and XGBoost has 90.5%. The advantage of RF lies in that it integrates the predictions of several decision trees. Such an approach will minimize overfitting and manage many features appropriately [12]. Thus, RF is an approach that is accurate and has a good effect. Organizations that experience customer churn prediction through RF are better placed when it comes to customer retention. It can enable businesses to do better in satisfying customers, thereby increasing more revenue and growth [6]. Even though other algorithms have good rates of accuracy as well, the RF works more effectively and significantly because they can handle heavy data. In this paper, RF is the main model to predict customer churn in e-Commerce because it demonstrates high accuracy and stability when dealing with complex and large amounts of data.

2.3.2 Support Vector Machine (SVM)

This paper is aimed at developing customer segmentation and customer churn prediction solutions to the B2C e-Commerce companies to help in coming up with effective customer retention initiatives and saving the operating expenses. Customer clustering was performed using the k-means type of algorithm and the LR and SVM algorithms to develop models of the prediction. The findings indicate that SVM is high compared to LR in relation to accuracy, recall, and AUC value. The average value of SVM classifier accuracy is 0.9156 and the average value of LR accuracy is 0.9066. The SVM model could locate and forecast almost 0.9721 recall and 0.8608 of precision and target at-risk customers more efficiently and correctly. SVM had AUC values of 0.992, 0.966 and 0.927 which were superior to that of LR model. Yet, the training time of SVM data is significantly shorter as compared to LR. According to experimental results, it was revealed that the SVM model has great generalization capacity and excellent prediction capacity. Thus, the SVM model will be a more realistic and efficient option in this instance as a means of speculating churning of the customers [13]. These observations indicate that SVM will find it easier to predict risks of customer churn and offer practical information. The predictive character of SVM shows its utility in the practical domain where it is important to make timely and accurate predictions to minimize the operating expenses and maximize customer retention. In this

paper, the SVM to predict customer churn in e-Commerce and verify its stability and accuracy on different datasets, thereby providing support for enterprises to formulate more effective customer retention strategies.

2.3.3 Gradient Boost (GB)

According to the paper, GB and LR algorithms are tested. The GB technique will be tested since the model is fairly accurate and fast in terms of processing [11]. The outcomes indicate that GB is better. In the training, it is 93% and in testing, it is 91% accurate. Comparatively, the LR training accuracy rate is 87% with testing accuracy rate of 88%. This implies that GB estimates better how to predict the churned customers. So, GB in the paper is more coherent than LR. The algorithm of the GB model could be applied as the estimation of precision comparison [14]. Nevertheless, despite its low level of precision, LR can also be applied in a model scenario, which involves simplicity and simple interpretation. The high level of accuracy of GB proves its benefits and exactly fits the case of tasks where accuracy is expected to be high and used practically. It would also be useful as a reference model for other similar studies. In this paper, GB and LR were compared on their ability to predict customer churn in e-Commerce, and it was found that GB was more accurate and better generalization so that it is more suitable to complex prediction tasks.

2.3.4 Logistic Regression (LR)

Based on the paper, six different ML models are studied and a new "Hybrid Customer churn Prediction (HCP)" model is constructed. The models involved include SVM, LR, RF, KNN, Multilayer Perceptron (MLP), GB, and hybrid churn prediction. However, HCP's dataset has an accuracy of 79.6 and 86 on Dataset 1 and Dataset 2, respectively. When studying two different datasets, the performance LR is different under different evaluation measures. For Dataset 1, the accuracy was 78.7% while the model of Dataset 2 improved the accuracy to 85%. The paper results illustrate the advantages and disadvantages of LR. It works very well in predicting customers who will not be churned. However, it is not effective in predicting churn customers, especially in the case of data imbalance. Therefore, we need to use more advanced models. It is also possible to combine several methods to improve the accuracy of prediction [15]. In this paper, it compared multiple ML models on two e-Commerce datasets and emphasized the effectiveness of hybrid methods in improving prediction accuracy, especially in handling data imbalance and enhancing customer churn detection. Although LR model simple and interpretable, often underperforms compared to ensemble models in real-world churn settings.

2.3.5 Decision Tree (DT)

The paper was dedicated to the forecast of customer churn in a new form of e-Commerce based on the information reflected in sales between customers and influencers. The paper has analysed the application of decision trees on analysis due to their high accuracy and ease of use. The analysis of the evaluation results indicates that the F-measure accuracy of this model is sufficient at 90% level which signifies the accurateness and significance of this model in customer churn prediction. The effective wasters were well detected by the model where the number of true positive (TP) was 71,591 and True Negative (TN) was 113. Despite low misclassifications, such as 15,570 False Positives (FP) and 146 False Negatives (FN), DT had a consistent performance. The decision trees can be proved as convenient and successful when solving the issue of prediction of customer churn in online business. They are accurate of 90% and a good choice, capable of providing a powerful instrument of identifying the at-risk customers [16]. Some of the classifications are erroneous, but this does not overshadow the general performance of the model, which points to the capacity of the decision tree in deriving usable insights. It is this that makes them highly suitable for businesses that require predictive models that are easy to use and are also highly accurate to boost customer retention and maintain a competitive edge. In this paper, decision tree models are also applied to predict customer churn in e-Commerce, and the results support their effectiveness as a simple and reliable tool for identifying customers at risk of leaving.

2.3.6 Summary and Gap

Overall, ensemble methods such as RF and GB perform better than simpler models because they handle complex data well and reduce overfitting. RF is accurate and stable, while GB works fast and generalizes well. SVM can identify at-

risk customers accurately, but it may need extra preprocessing. DT and LR are easy to use and interpret but often perform worse on complex or imbalanced data. Most studies focus only on accuracy with static datasets. They often ignore real-time prediction, computational cost, scalability, and business integration. Ensemble models are accurate but can be hard to interpret for business use. Future research should combine strong algorithms with explainable AI, real-time data, and cost analysis. This will help businesses use churn prediction models that are both accurate and practical.

2.4 Comparison between Machine Learning Methods

In this part, there are several papers that are presented with the focus on outlining the paper's objective, ML algorithms to be applied, the results, and the contributions of the studies. The paper by [4] published a paper that intended to make effective ML models to predict the customer churn in e-Commerce. In the paper, there were several algorithms, such as LLR, LDA, DT, KNN, XGBoost and RF. The findings reveal that the accuracy rate of LDA and DT equals to 96%, whereas the accuracy of the RF equals to 98%. With this paper, it can be concluded that the discriminant model and the ensemble model would do great in predicting customer churn. It can also be used to refer to the comparisons between the performances of several algorithms on the same dataset.

The paper of [14] utilized behavioural data from Alibaba Cloud Tianchi. The data contains nearly one million user records. They applied LR and SVM for prediction. The average accuracy rate of LR is 90.66%, and that of SVM is 91.56%. Before modelling, they used K-means clustering for customer segmentation. This improves the accuracy of the model. Paper show that the performance of SVM is superior to that of traditional models. This paper provides B2C e-Commerce platforms with personalized retention strategies based on behavioural patterns.

[15] conducted a paper using a dataset of 5,630 records. They compared the predictive capabilities of LR and GB. The test accuracy rate of LR is 93%, and that of GB is 91%. This paper shows that tree models and statistical models are very effective in customer churn prediction. The results also prove that GB performs well in the training and testing phases. This paper demonstrates the application potential of a simple yet powerful model in actual business, especially suitable for small and medium-sized datasets.

[16] aims at forecasting customer churn within influence-driven e-Commerce platforms. In this paper, sales information of one of marketing firms of South Korea was used. This was a modelling study that was done using the DT algorithm. The recall rate of the model is 99.7%, the precision rate 82.1%, the accuracy rate 82% and the F-measure 90%. The paper indicate that DT are suitable for predicting customer churn. Some of the main factors that could affect the behaviour of customers were also noted in the research and recommendations could also be given to enterprises so that they could develop customer retention strategies.

[17] proposed a ML method to predict customer churn. Four algorithms are test which are LR, DT, RF and AdaBoost. The accuracy rate of RF has reached 96%, and that of AdaBoost is 89%. This paper combines exploratory data analysis and classification techniques. It helps e-Commerce enterprises better understand customer behaviour. The paper provides an effective method for implementing proactive customer retention strategies.

The paper by [10] prediction of customer churn, annual spending and on-time delivery using ML algorithms. The paper used three different datasets from Kaggle. For customer churn prediction, in Dataset 1, the accuracy of SVM is the highest, reaching 83.45%. In Dataset 2 uses linear regression model, the accuracy rate of the annual spending forecast is 97.82%. In terms of delivery prediction in dataset 3, the scores of both SVM and decision tree are about 68%. This paper combines predictive models with consumer behaviour, logistics and financial forecasts. It provides comprehensive decision-making support for e-Commerce.

[18] conducted a paper using a data from HoReCa and the retail industry. The paper compared multiple ML models to predict purchasing behaviour. The models used include LR, DT, KNN, Naive Bayes, SVM, RF, neural network, AdaBoost and XGBoost. The accuracy rates of both RF and XGBoost have reached 91.67%. The accuracy of the hybrid model KnnSgd is higher, reaching 92.42%. This paper demonstrates the advantages of model combinations and evaluates them using cross-validation and confusion matrices. The paper provides a reliable method for practical applications.

[19] applied models such as LR, DT, RF, XGBoost and LightGBM to predict customer churn. The accuracy of XGBoost is the highest, reaching 96%, while that of LightGBM is 93%. This paper uses the SMOTE method to handle the problem of class imbalance. The paper also used SHapley Additive exPlanations (SHAP) and Local Interpretable

Model-agnostic Explanations (LIME) to interpret the model results. This paper helps enterprises understand the key factors of customer churn. It provides support for formulating targeted retention strategies.

[20] conducted a paper to predict of customer churn in e-Commerce. This paper used six algorithms which are LR, DT, RF SVM, XGBoost and AdaBoost. The dataset contains behavioural and demographic characteristics. The accuracy rate of RF is the highest, at 96.81%, followed closely by XGBoost, at 96.27%. This paper made a detailed comparison of the performances of various algorithms. It emphasizes the advantages of the integrated approach. At the same time, the overfitting problem was also dealt with. The paper results are applicable to the actual customer churn management.

Lastly, [21] focused predict customer churn. They evaluated seven ML models, including ANN, SVM, KNN and ensemble methods. ANN has the highest accuracy rate, which is 92.09%, followed by RF, which is 91.21%. This paper combined interpretable AI tools such as SHAP and LIME to enhance the transparency of the model. The paper provides guidance for choosing the appropriate model and interpreting the results. It is very helpful for practitioners who hope to use explainable AI to improve customer retention rates.

Table 1 shows the summary table, which presents the different ML techniques that are applied in predicting customer churn in e-Commerce. The result showed that ensemble models performed better, especially RF and XGBoost that obtained high accuracy in most of the papers such as the papers of [6], [17] and [20]. These models are particularly effective because they can handle complex data patterns and reduce the problem of overfitting. Moreover, certain papers like [10] and [19] applied SMOTE to solve the issue of dataset class imbalance, which considerably increased the predictive capability of the models that were used in their paper. Besides, the interpretability of models has become more popular in recent papers.

Table 1. Summary Table

Reference	**Objective**	**Dataset**	**ML Algorithm**	**Result (%)**	**Value added**
[21]	Compare model performance for churn prediction and explore interpretability	Kaggle dataset	DT RF NB LR SVM KNN ANN	DT = 89.81 RF = 91.21 NB = 75.04 LR = 85.59 SVM = 83.3 KNN = 84.18 ANN = 92.09	Applies XAI technology such as SHAP and LIME to enhance the interpretability of the model and provides business-oriented practical suggestions for optimizing customer churn strategies.
[10]	ML to predict churn, spending, and delivery success	Kaggle (3 datasets)	Linear Regression DT SVM RF XGBoost	Dataset 1: XGBoost= 83.38 DT=83.38 RF=82.10 SVM=83.45 Dataset2: Linear regression =97.82 Dataset 3: XGBoost =67.09 DT=68.48 RF=66.93 SVM=68.42	Used multiple models for different predictive targets, to support customer insight analysis and the formulation of retention strategies.
[19]	Identify churn factors and provide retention recommendations	Kaggle dataset	LR DT RF XGBoost LightGBM	LR = 78 DT = 92 RF = 86 XGBoost = 96 LightGBM = 93	Use SMOTE to balance the data and use SHAP and LIME for interpretation. Link the model output with the actual situation and adopt targeted retention strategies to reduce attrition and enhance profitability.

[6]	Build effective ML models to predict customer churn in e-Commerce	Kaggle dataset with e-Commerce behavior	LR LDA DT KNN RF XGBoost	LR = 85 LDA = 96 DT = 96 KNN = 91 RF = 98 XGBoost = 90.5	Comparative analysis of six ML algorithms. Used SMOTE to handle imbalance. Found RF achieved highest accuracy. Provided practical insights to support churn reduction strategies.
[17]	Apply ML for churn prediction to drive engagement	Kaggle dataset	LR DT RF AdaBoost	LR = 88 DT = 87 RF = 96 AdaBoost = 89	This paper verified the effectiveness of the model through EDA and classification methods and provided profound insights focused on enhancing customer retention rates.
[18]	Predict customer purchasing behaviour using ML classification techniques	UCI Machine Learning Repository	LR DT KNN RF XGBoost KNN-SGD	LR = 89.29 DT = 89.29 KNN = 90.48 RF = 91.67 XGBoost = 91.67 KNN-SGD = 92.42	Compares multiple supervised ML and ensemble methods using cross-validation, confusion matrices, and ROC analysis. The hybrid KNN-SGD classifier achieved the highest accuracy of 92.42%.
[13]	Predict churn via shopping behavior and enhance retention	Alibaba Tianchi dataset	LR SVM	LR = 90.66 SVM = 91.56	This paper combines clustering with predictive models, with SVM performing the best, providing a data-driven and precise target positioning solution for B2C e-Commerce retention strategies.
[14]	Enable effective ML use for churn prediction	Kaggle dataset	LR Gradient Boost	Training Accuracy LR = 87 GB = 88 Testing Accuracy: LR = 93 GB = 91	Compared two models for churn prediction. Showed Gradient Boosting had better performance. Validated the usefulness of ML techniques and data mining for customer churn analysis.
[16]	Predict customer churn in influencer-driven e-Commerce	Influencer sales data from a Korean marketing agency	Decision Tree	Recall = 99.7 Precision = 82.1 Accuracy = 82 F-Measure = 90	Achieved high recall and feasibility using DT. Identified key churn drivers. Provided recommendations for improving retention strategies. Established foundation for future research using more advanced models.
[20]	Predict churn and identify key behavioral and demographic churn factors	E-Commerce dataset	LR DT RF SVM XGBoost AdaBoost	LR = 77.35 DT = 93.47 RF = 96.81 SVM = 88.07 XGBoost = 96.27 AdaBoost = 85.20	A comprehensive evaluation of six models, emphasizing random forests has the best model and highlighting the significance of features and the interpretability of the models.

In the studies conducted in [2] and [19], SHAP and LIME technologies were adopted to provide clarity on model predictions and explain customer behaviour leading to churn. These interpretability techniques show which features most influence the model's decisions, helping stakeholders understand why certain customers are predicted to churn. This is particularly useful for complex models like Random Forest, which are often considered "black boxes." Furthermore, most papers have employed comparisons of several algorithms to determine the robustness and

generalization of the results. Some papers have also utilized clustering techniques to enhance customer segmentation and customer retention strategies. Although most papers used the publicly available datasets of Kaggle, there were also some studies like [13] and [16] that utilized different data sources, such as Alibaba Tianchi or proprietary datasets. These demonstrate that ML models have good adaptability and flexibility in various types of data environments. Overall, these papers indicate that the complexity of high-performance ensemble algorithms, data preprocessing methods, and interpretability tools is expected to provide a solid foundation for building a temporarily sufficient customer churn prediction system and propose sensible, data-based customer retention policies.

3. METHODOLOGY

3.1 Proposed Framework

The purpose of this framework is to develop an ML-based framework for predicting customer churn in e-Commerce. It identifies at-risk customers by analysing historical behavioural data, such as purchase patterns, transaction frequencies and customer demographics. By using this framework, e-Commerce companies can proactively identify churn customers and implement retention strategies, thereby reducing customer churn and increasing customer retention rates.

Figure 1 illustrates the proposed framework, showing the sequence of steps from data collection to model evaluation and interpretation. At the first step of this paper, preparational work of the dataset was performed first. Compared to most of the previous research study, which only utilized a single set of data, this paper employed the use of two sets of data related to the e-Commerce business. It also enables us to test the generalization ability of the created model in an alternating data environment, which demonstrates the innovation and practicality of the current research method and thus makes the research more powerful. Thus, in terms of customer churn prediction, only customer behaviour related data including customer interaction, customer purchase history, customer reviews and others are gathered.

The next step was data preprocessing, which involved several stages: data cleaning, feature selection or engineering, segmentation, and handling of imbalanced classes. Missing values were imputed using mean, median, or mode, depending on the feature type. Categorical variables were encoded using one-hot encoding or label encoding as appropriate. Feature selection techniques were applied to remove low-variance or irrelevant features, and important variables were retained to improve model performance.

Following preprocessing, five ML algorithms which are DT, RF, LR, SVM, and GB were trained. These models were selected for their effectiveness in classification, interpretability, and ability to handle complex data patterns. All ML models were trained using their default hyperparameters in scikit-learn, with a few specific settings to ensure reproducibility and convergence. For Random Forest, Decision Tree, Gradient Boosting, and SVM, random_state=42 was set to guarantee consistent results across runs. Logistic Regression was trained with max_iter=1000 to ensure convergence on the dataset. Additionally, probability=True was enabled for SVM to allow probability predictions. Other parameters, such as the number of trees, maximum depth, and learning rate, were kept at their default values. This approach ensures that the models can be reproduced while maintaining simplicity in training.

Model performance was evaluated using standard metrics including accuracy, precision, recall, and F1 score. A decision checking mechanism was incorporated to verify predictions and improve reliability. Comparisons of model performance across the two datasets were conducted to assess stability and generalization. This framework ensures that all preprocessing steps, feature selection methods, and model parameters are clearly detailed, enhancing reproducibility and providing a solid foundation for future work, such as testing hybrid models or additional datasets.

3.2 Dataset

The dataset used in this study is from Kaggle. Dataset 1 contains 20 attributes and 3333 data. This dataset mainly records the detailed of user behaviours and interaction activities on the e-Commerce platform. The data includes attributes related to account information, browsing activities, purchasing behaviours and interaction records of users on the platform [22]. Table 2 shows the attributes and data descriptions contained in Dataset 1.

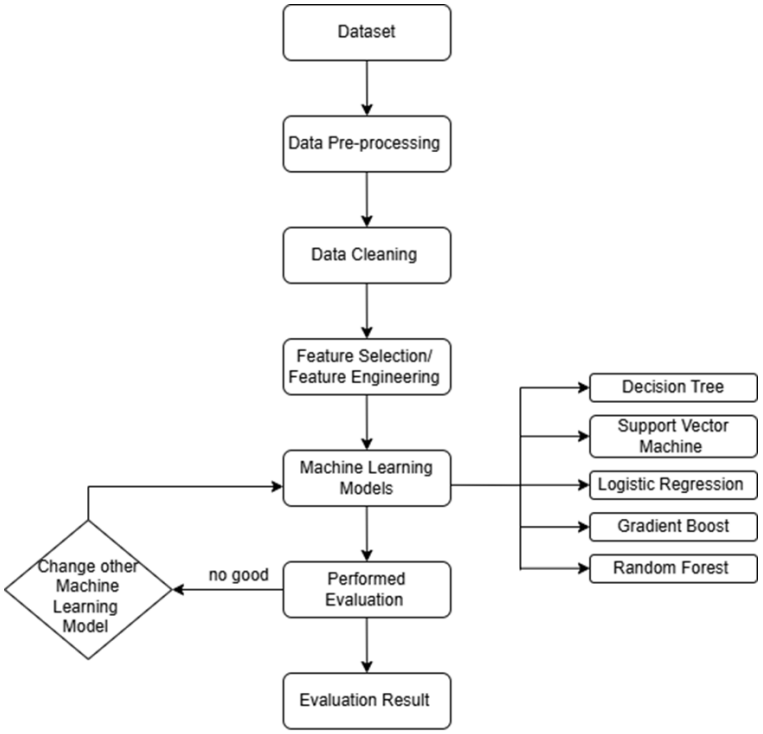


Figure 1. Propose Framework

Table 2. Attributes of Dataset 1

No.	**Attributes**	**Description**
1	account length	The days that user has an account on the platform.
2	location code	The region code of the user.
3	user id	Unique customer ID
4	credit card info save	Has the user has saved their credit card information on the platform (Yes/No).
5	push status	Has the user enabled push notifications? (Yes/No).
6	add to wishlist	The number of products that the user has added to their wish list.
7	desktop sessions	The number of times the user accesses the platform via desktop.
8	app sessions	The number of times the user accesses the platform via mobile app.
9	desktop transactions	The number of purchases was completed through the desktop platform
10	total product detail views	The total number of product information pages viewed by the user
11	session duration	The total time the user spent in sessions on the platform.
12	promotion clicks	Number of promotional offers clicked by the user.
13	avg order value	Average value of the user's orders.
14	sale product views	Number of sale or discounted products viewed by the user.
15	discount rate per visited products	Browse the average discount percentage of the product.
16	product detail view per app session	Average number of product detail pages viewed in each app session.
17	app transactions	The quantity purchased through the mobile app.
18	add to cart per session	The average number of products added to the cart for each session
19	customer service calls	The number of times the user contacts customer service.
20	churn	Has the customer churned? (Yes/No).

The Dataset 2 consists of 19 attributes and 5630 data. This dataset comes from a well-known online retailer and is designed to identify and predict the customers most likely to leave. This dataset includes various characteristics that reflect customers' interactions with the platform [23]. Table 3 shows the attributes and data descriptions contained in the Dataset 2.

Table 3. Attributes in Dataset 2

No.	Attributes	Description
1	Customer ID	Unique customer ID
2	Churn	Has the customer churned? (Yes/No).
3	Tenure	The tenue of the client within the organization
4	CityTier	CityTier
5	WarehouseToHome	The distance from the warehouse to the customer's home
6	PreferredPaymentMode	The preferred payment method of customer
7	Gender	Customer Gender
8	HourSpendOnApp	The time spent on mobile applications or websites
9	NumberOfDeviceRegistered	Total number of deceives is registered on particular customer
10	PreferedOrderCat	The customer's preferred order category last month
11	SatisfactionScore	The degree of customer satisfaction with the service
12	MaritalStatus	Customer's marital status
13	NumberOfAddress	Total number of added on particular customer
14	Complain	Any complaints were raised last month
15	OrderAmountHikeFromlastYear	Percentage increases in order from last year
16	CouponUsed	Total number of coupons has been used in last month
17	OrderCount	Total number of orders has been places in last month
18	DaySinceLastOrder	Day Since last order by customer
19	CashbackAmount	Average cashback in last month

3.3 Data Pre-Processing

This paper uses Python to handle missing values. After checking each dataset, Dataset 1 contains 20 attributes with no missing values and no duplicate data. However, in Dataset 2 it contains missing values in several columns such as Tenure 264 records, WarehouseToHome 251 records, HourSpendOnApp 255 records, OrderAmountHikeFromLastYear 265 records, CouponUsed 256 records, OrderCount 258 records and DaySinceLastOrder 307 records. To address these missing values, the missing value column for the Tenure, Coupon used, Day SinceLastOrder and HourSpendOnApp columns will fill up by using the media to ensure that the dataset remains balanced. The columns WarehouseToHome, OrderAmountHikeFromlastYear and OrderCount will use mean to keep the balance. After preprocessing, there are no missing values in the dataset. This can ensure that the data can be reliably used for analysis and modelling. Other columns, such as CustomerID, Churn, PreferredLoginDevice, and SatisfactionScore and others have no missing values. This dataset also does not contain duplicate data. Duplicate records were also checked and removed. Table 4 shows the distribution of missing values for each attribute in Dataset 2.

3.4 Feature Selection or Feature Engineering

In Dataset 1, categorical variables such as credit card info save, and push status are coded using label coding. Label encoding converts each category into a unique integer, allowing these columns to be used as input features for ML algorithms. By converting these classification features to numerical values, the data set is better suited for predictive modelling to allow the algorithm to process the data more efficiently. The encoding and decoding process for these categorical values in Dataset 1 is illustrated in Table 5.

Table 4. Missing Value in Dataset 2

Attribute	Missing Values
CustomerID	0
Churn	0
Tenure	264
PreferredLoginDevice	0
CityTier	0
WarehouseToHome	251
PreferredPaymentMode	0
Gender	0
HourSpendOnApp	255
NumberOfDeviceRegistered	0
PreferedOrderCat	0
SatisfactionScore	0
MaritalStatus	0
NumberOfAddress	0
Complain	0
OrderAmountHikeFromlastYear	265
CouponUsed	256
OrderCount	258
DaySinceLastOrder	307
CashbackAmount	0

Table 5 (a). Dataset 1 After Feature Engineering

Account Length	Location Code	Credit Card Info Save	Push Status	Add to Wishlist	Desktop Sessions	App Sessions	Desktop Transactions	Total Product Detail Views	Session Duration
0	128	415	0	1	25	265	45	17	197
1	107	415	0	1	26	162	27	17	196
2	137	415	0	0	0	243	41	10	121
3	84	408	1	0	0	299	51	5	62
4	75	415	1	0	0	167	28	13	148
…	…	…	…	…	…	…	…	…	…
3328	192	415	0	1	36	156	27	18	216
3329	68	415	0	0	231	39	13	57	153
3330	28	510	0	0	181	31	25	109	289
3331	184	510	0	0	214	36	14	105	160
3332	74	415	0	1	25	234	40	23	266

In Dataset 2, categoric variables such as PreferredLoginDevice, PreferredPaymentMode, Gender, PreferedOrderCat, and MaritalStatus into numerical form. These features are converted into data formats that ML models can understand. By converting each category to a number, label encoding enhances the applicability of the dataset to analysis and model training. This process simplifies the data representation by converting non-digital data into digital form and improves the performance and predictive ability of the model. The encoding and decoding process for these categorical variables in Dataset 2 is illustrated in Table 6.

Table 5 (b). Dataset 1 After Feature Engineering

Promotion Clicks	Avg Order Value	Sale Product Views	Discount Rate per Visited Products	Product Detail View per App Session	App Transactions	Add to Cart per Session	Customer Service Calls	Churn
87	244.7	91	11.01	10.0	3	2.70	1	0
103	254.4	103	11.45	13.7	3	3.70	1	0
110	162.6	104	7.32	12.2	5	3.29	0	0
88	196.9	89	8.86	6.6	7	1.78	2	0
122	186.9	121	8.41	10.1	3	2.73	3	0
126	279.1	83	12.56	9.9	6	2.67	2	0
55	191.3	123	8.61	9.6	4	2.59	3	0
58	191.9	91	8.64	14.1	6	3.81	2	0
84	139.2	137	6.26	5.0	10	1.35	2	0
82	241.4	77	10.86	13.7	4	3.70	0	0

Table 6 (a). Dataset 2 After Feature Engineering

Churn	Tenure	Preferred Login Device	City Tier	Warehouse To Home	Preferred Payment Mode	Gender	Hour Spend On App	Number Of Device Registered	Preferred Order Category
0	1	4.0	1	3	6.0	4	0	3.0	2
1	1	9.0	2	1	8.0	6	1	3.0	3
2	1	9.0	2	1	30.0	4	1	2.0	3
3	1	0.0	2	3	15.0	4	1	2.0	2
4	1	0.0	2	1	12.0	0	1	3.0	3
5	1	9.0	2	1	22.0	0	1	3.0	4
6	1	9.0	2	3	11.0	2	1	3.0	2
7	1	9.0	1	1	6.0	0	1	3.0	3
8	1	13.0	2	3	9.0	0	1	3.0	3
9	1	9.0	2	1	31.0	4	1	2.0	3

4. IMPLEMENTATION

4.1 Machine Learning Model Evaluation

To predict customer churn, several ML models will be developed and evaluated. The objective is to find the best model to reliably identify which customer have leave. Train and test each model using the same dataset and evaluate the performance of each model using key assessment metrics such as accuracy, precision, recall, and F1 score. This study employs LR, DT, SVM, RF and GB models. When dealing with data complexity, bias and variance, each algorithm has its unique advantages and disadvantages. 70% of each dataset is used for training and 30% for testing. The results and discussions of each model will be presented in the following sections.

Table 6 (b). Dataset 2 After Feature Engineering

Satisfaction Score	Marital Status	Number Of Address	Complain	Order Amount Like From Last Year	Coupon Used	Order Count	Days Since Last Order	Cashback Amount	Tenure Range
2	2	9	1	11.0	1.0	1.0	5.0	160	0–18
3	2	7	1	15.0	1.0	1.0	0.0	121	0–18
3	2	6	1	14.0	0.0	1.0	3.0	120	0–18
5	2	8	0	23.0	0.0	1.0	3.0	134	0–18
5	2	3	0	11.0	1.0	1.0	3.0	130	0–18
5	2	2	1	14.0	4.0	6.0	0.0	139	0–18
2	0	4	0	14.0	0.0	1.0	0.0	121	0–18
2	0	3	1	16.0	2.0	2.0	0.0	123	0–18
3	2	2	0	12.0	1.0	1.0	1.0	123	0–18

4.1.1 Random Forest

In Dataset 1, the confusion matrix of the RF classifier, shown in Figure 2, illustrates the segmentation of true positives, true negatives, false positives, and false negatives. Among the 857 non-churn customers, 840 were accurately predicted, and 17 were incorrectly classified as churn customers. For the churn class, 114 customers were correctly identified, while 29 were wrongly predicted as non-churn customers. This indicates that the model is highly effective in predicting non-churn customers, but its ability to accurately identify churn customers is comparatively lower. The relatively high number of false negatives (29) indicates that this model sometimes fails to detect churn cases, which is critical for business strategies focused on customer retention.

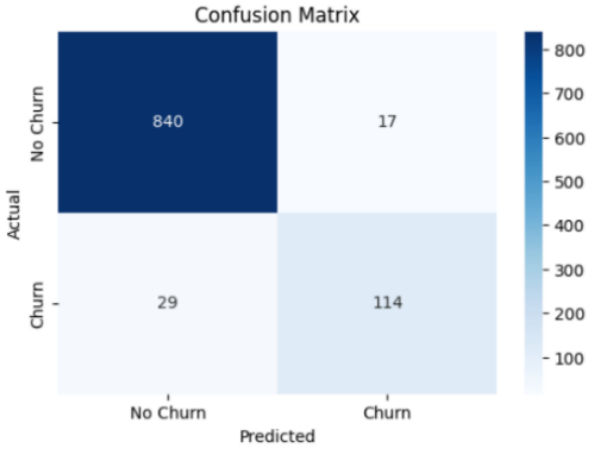


Figure 2. Confusion Matrix of RF Classifier in Dataset 1

In Dataset 2, the confusion matrix of the RF classifier, shown in Figure 3, illustrates the model's performance in predicting customer churn. Among 1,414 actual non-churn rate customers, 1,397 were correctly identified, while 17 incorrectly classified as churn. For the churn category, among 275 customers, the model correctly predicted 231 churn cases but wrongly classified 44 as non-churn. This distribution highlights that the model performs very well in predicting non-churn customers, which is reflected in the low number of false positives (17). Nevertheless, a considerable number of churn cases (44) have been overlooked, which may affect the efforts to proactively retain customers. Although this model demonstrates strong overall predictive capabilities, especially for non-churn customers. However, to enhance its sensitivity to churn cases will further improve its practical effectiveness in customer retention strategies.

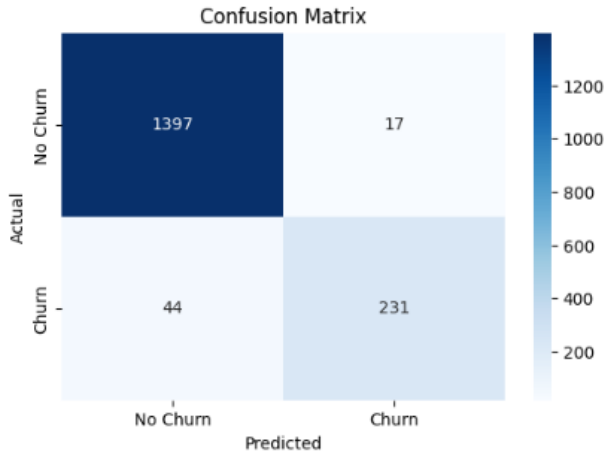


Figure 3. Confusion Matrix of RF Classifier in Dataset 2

4.1.2 Logistic Regression

In Dataset 1, the confusion matrix of the LR classifier, shown in Figure 4, illustrates the model's performance in predicting customer churn. Among the 857 actual non-churn customers, 832 were correctly identified, while 25 were wrongly classified as churn customers. For churn customers, only 31 out of 143 customers were accurately predicted, while 112 were wrongly classified as non-churn customers. These results indicate that the model is highly effective in identifying potential customers who may remain, but it is very difficult in identifying potential customers who may leave. Many false negatives (112) highlight a major weakness of many churn customers who have been overlooked, which may pose a serious challenge to the efforts to retain employees. Although the accuracy of this model is still quite high, enhancing its ability to detect lost customers is crucial for more effective business decisions.

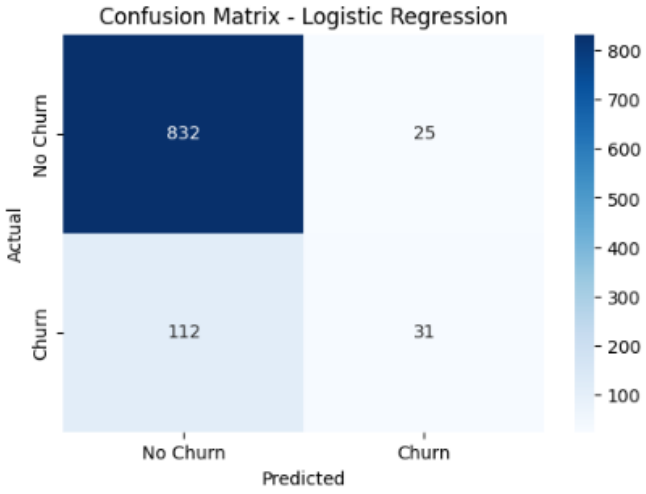


Figure 4. Confusion Matrix of LR Classifier in Dataset 1

In Dataset 2, the confusion matrix of the LR classifier, shown in Figure 5, illustrates the model's performance in predicting customer churn. Among the 1,414 actual non-churn customers, 1,371 were correctly identified, while 43 were wrongly classified as churn customers. For 275 churn customers, the model accurately predicted 123 cases, but 152 churn customers were wrongly classified as non-churn customers. This indicates that the model is effective in identifying non-churn customers and has high accuracy and recall rate for the non-churn category. However, it is difficult to accurately predict the churn customers, which can be evidenced by the relatively high number of false negatives (152). It is crucial to enhance the model's ability to predict customer churn, as it is currently missing out on many customers at risk of leaving.

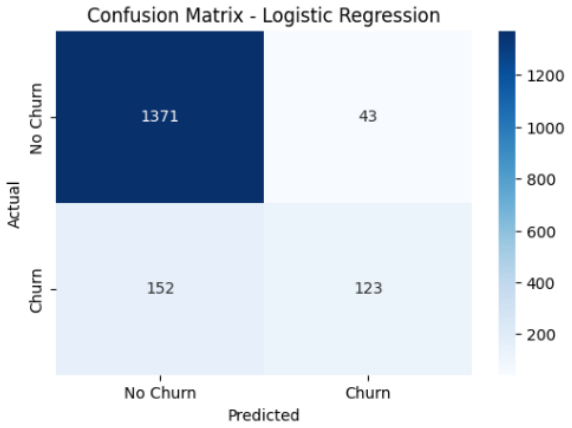


Figure 5. Confusion Matrix of LR Classifier in Dataset 2

4.1.3 Decision Tree

In Dataset 1, the confusion matrix of the DT classifier, shown in Figure 6, illustrates the model's performance in predicting customer churn. Among the 857 non-churn rate customers, 64 were wrongly classified as churn rate customers, while 793 customers were correctly identified as non-churn rate. This model accurately predicted 104 out of 143 churn customers, but 39 churn customers were wrongly marked as non-churn customers. The results of false positives and false negatives indicate that the model performs well in identifying non-churn customers, but poorly in predicting churn customers.

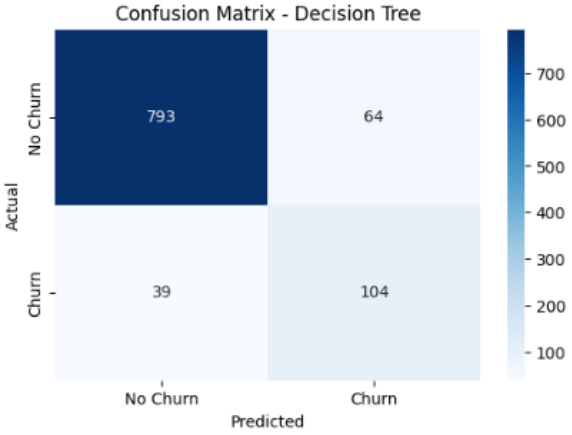


Figure 6. Confusion Matrix of DT Classifier in Dataset 1

In Dataset 2, the confusion matrix of the DT classifier, shown in Figure 7, illustrates the model's performance. Among 1,414 non-churn customers, 1,374 were correctly predicted as non-churn, and only 40 were wrongly predicted as churn customers. Among the 275 actual churn customers, 240 were correctly identified, while only 35 were wrongly regarded as non-churn customers. This indicates that the misclassification rate of this model is very low, and it performs well in predicting both churn and non-churn customers. Overall, the model performed well, with a relatively small number of false positives and false negatives.

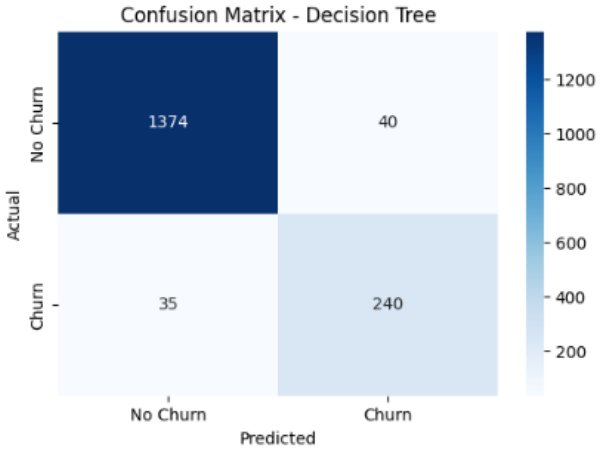


Figure 7. Confusion Matrix of DT Classifier in Dataset 2

4.1.4 Support Vector Machines

In Dataset 1, the confusion matrix of the SVM classifier, shown in Figure 8, illustrates its performance. The model correctly identified 849 non-churn customers (true negatives) and 70 churn customers (true positives). However, it misclassified 73 actual churn customers as non-churn (false negatives) and 8 non-churn customers as churn (false positives). This indicates that the model are accurately predicts non-churn customers but has a relatively weak detection ability for churn customers. There are many false negatives, indicating that the model has missed a considerable number of churn customers. Besides, improving the model's ability to identify churn customers is of great significance for enhancing overall performance, especially when the prediction of churn customers has a significant impact on business.

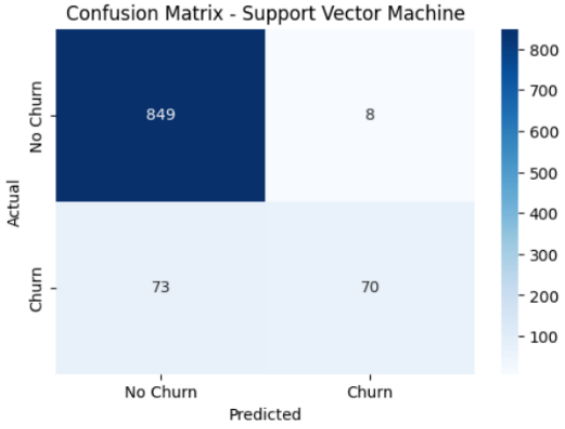


Figure 8. Confusion Matrix of SVM Classifier in Dataset 1

In Dataset 2, the confusion matrix of the SVM classifier, shown in Figure 9, illustrates its performance. The model correctly identified 1,398 non-churn customers and 136 churn customers. However, this model wrongly identified 139 churn customers as non-churn customers, and only 16 non-churn customers were wrongly identified as churn customers. Although the overall performance of this model is relatively good and there are many correct judgments, the number of false negatives is relatively large, indicating that this model has missed many churn customers.

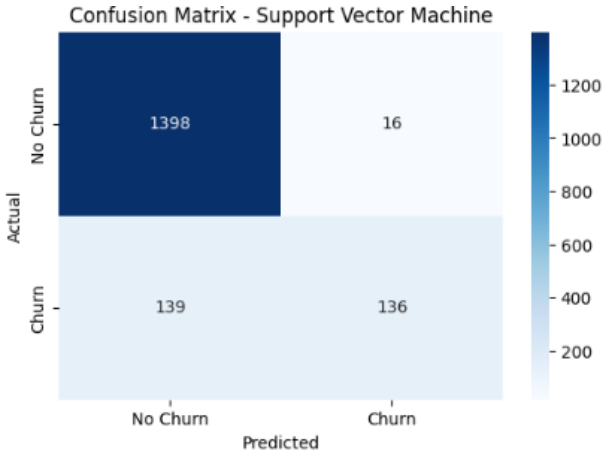


Figure 9. Confusion Matrix of SVM Classifier in Dataset 2

4.1.5 Gradient Boost

In Dataset 1, the overall performance of the GB classifier is good, as shown in the confusion matrix in Figure 10. It managed to recognize 852 non-churn customers and 102 churn customers. The model failed also to predict 41 actual churn customers. In addition, 5 non-churn customers were wrongly predicted as churn customers. Although the number of false alarms was very small with only 5, the number of lost customers missed was relatively large with 41. This indicates that the model is very powerful in identifying non-churn customers, but there is still room for improvement in identifying churn customers. To improve the prediction of customer churn, the model is to be optimized, and it should increase the recall rate. This can assist businesses to identify and retain potential customers who may be churn.

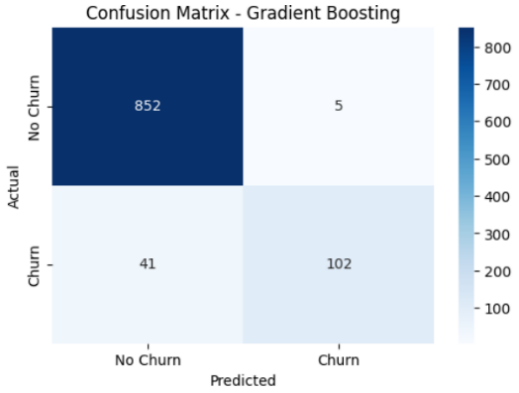


Figure 10. Confusion Matrix of GB Classifier in Dataset 1

In Dataset 2, the confusion matrix of GB classifier, shown in Figure 11, indicates that the model performs well in identifying non-churn customers, but there still have change to improvement in detecting churn customers. The model successfully identified 1,373 non-churn customers and 168 churn customers. However, this model also wrongly identified 107 actual churn customers as non-churn customers and wrongly classified 41 non-churn customers as churn customers. Although the number of false positives is relatively small, the large number of false negatives indicates that this model still has obvious deficiencies when identifying churn customers. Overall, the model performs well in terms of overall classification. However, if the ability to identify churn customers can be enhanced, it will be more reliable and practical in actual business operations.

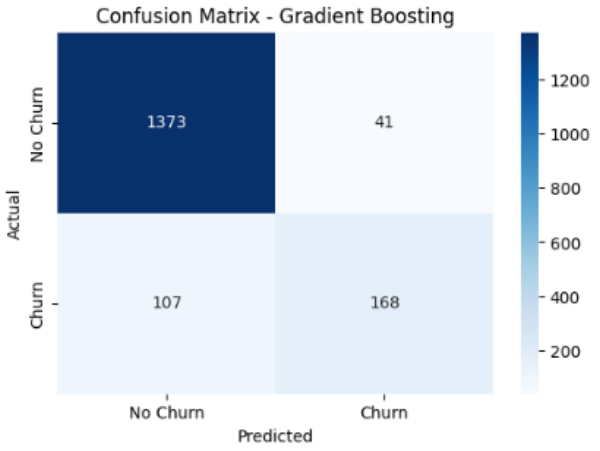


Figure 11. Confusion Matrix of GB Classifier in Dataset 2

5. RESULTS AND DISCUSSIONS

5.1 Model Performance Comparison

This section compares the performance of five ML models on two different datasets, including LR, DT, SVM, GB and RF. The evaluation indicators of the model include accuracy, precision, recall and F1 score. In Dataset 1, the RF performed the best with an accuracy of 0.95, and the accuracy of the best F1 score was 0.83, indicating that it is very effective in identifying customer churn. However, GB also performed well, with an accuracy of 0.94 and an F1 score of 0.78. The performance of the DT is also relatively stable, with a high recall, and it can well identify the churn customers. In contrast, the recall rate of LR is very low, and it can only identify a small number of churn customers. Therefore, the overall performance of LR is poor. Besides, although SVM has a high accuracy, its low recall indicates that it often misses some churn customers. This low recall across several models further highlights the impact of data imbalance on churn prediction, where minority churn cases are often misclassified. Addressing this issue with SMOTE or cost-sensitive learning can potentially improve recall performance without sacrificing accuracy.

In Dataset 2, RF and DT perform the best. Their accuracy has all reached 0.96. The F1 score of the RF is 0.88, indicating that it can well identify whether customers are churn. The F1 score of the DT is 0.86 and it also performs very stably. In contrast, the effects of SVM and GB models are average. Their recall rates were relatively low, at 0.49 and 0.61 respectively, indicating that they were not easy to identify churn customers.

In conclusion, the results of these two datasets indicate that the RF model has achieved the highest accuracy in both datasets. In Dataset 1, RF achieved the highest accuracy (0.95) and F1-score (0.83). In Dataset 2, RF and DT achieved equal accuracy (0.96), but RF obtained a higher F1-score (0.88) and precision (0.93), while DT achieved higher recall (0.87). Overall, RF demonstrated stronger and more balanced performance for customer churn prediction. Therefore, when businesses conduct churn prediction, they can give priority to using the RF model. Table 7 and Table 8 shown the contrast between performance metrics of ML models in Dataset 1 and Dataset 2. To complement these quantitative results, the SHAP and LIME analyses can be used to interpret model predictions, helping businesses understand which factors such as tenure, frequency of transactions, or customer complaints are most influential in driving churn. Such interpretability strengthens the practical value of the models for decision-making in customer retention.

Table 7. Contrast between Performance Metrics of ML Models in Dataset 1

Model	**Comparison**			
	Accuracy	**Precision**	**Recall**	**F1-Score**
Random Forest	0.95	0.87	0.80	0.83
Gradient Boosting	0.94	0.78	0.78	0.78
SVM	0.92	0.90	0.49	0.63
Decision Tree	0.90	0.62	0.73	0.67
Logistic Regression	0.86	0.55	0.22	0.31

Table 8. Contrast between Performance Metrics of ML Models in Dataset 2

Model	**Comparison**			
	Accuracy	**Precision**	**Recall**	**F1-Score**
Random Forest	0.96	0.93	0.84	0.88
Decision Tree	0.96	0.86	0.87	0.86
SVM	0.91	0.89	0.49	0.64
Gradient Boosting	0.91	0.80	0.49	0.69
Logistic Regression	0.88	0.74	0.61	0.56

5.2 Discussion on Results

The model accuracy for Dataset 1 and Dataset 2 is compared in Figure 12 and Figure 13 respectively. The results from both datasets show that RF gave the best performance. In Dataset 1, RF reached an accuracy of 0.95 and an F1 score of 0.83. In Dataset 2, it reached 0.96 accuracy and an F1 score of 0.88. This supports the findings in Section 2.3.1, which showed that ensemble methods usually work better than single models because they combine many decision trees and reduce errors. RF works well because it can capture complex and nonlinear customer behaviour, which is common in churn data.

LR and SVM performed worse. LR had the lowest results in Dataset 1 with accuracy of 0.86 and F1 score of 0.31. This shows that LR has problems when the data is nonlinear or imbalanced. SVM had high accuracy (0.92) but very low recall (0.49). This means it often failed to detect customers who actually churned. Previous studies also showed that SVM struggles when the data is imbalanced. In practice, this is risky because missing churners (false negatives) can cause businesses to lose customers and revenue.

Dataset 2 gave similar results. Both RF and DT performed very well, with accuracy of 0.96 and F1 scores of 0.88 and 0.86. DT is easier to interpret, which makes it useful for managers, even though it is weaker than RF in predictive power. RF, however, is both strong and adaptable, which explains its success across different datasets. On the other hand, SVM and Gradient Boosting had moderate recall (0.49–0.61), showing again that accuracy alone is not enough to judge model quality.

These results have clear business meaning. Companies should not focus only on accuracy. Previous research also shows that missing churners is more costly than wrongly predicting churn for loyal customers. This makes recall and F1 score more important. Also, even though RF is powerful, interpretability matters. Tools like SHAP or LIME can be used with RF to show why customers churn, which makes the results easier for businesses to apply.

The weak performance of LR and SVM also shows that model choice depends on the data. LR is not good for nonlinear and complex features. SVM performs poorly with imbalanced data. If companies use these models without adjustments, they may waste resources and fail to keep customers. Techniques like cost-sensitive learning, oversampling, or hybrid models may be needed to improve results.

In conclusion, ensemble models, especially RF, are very effective for churn prediction. But two challenges remain. First, class imbalance must be handled, because accuracy can hide poor performance. Second, models should not only predict well but also give explanations that managers can use. Future work should test hybrid or cost-sensitive models that improve recall while keeping the results easy to understand and apply.

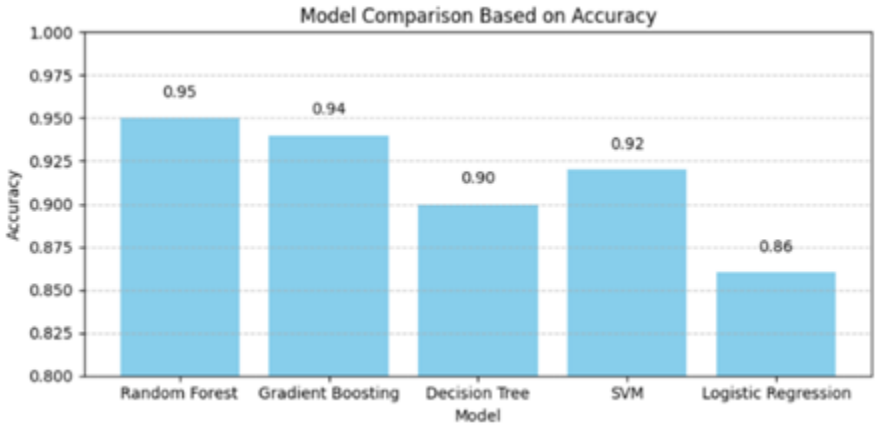


Figure 12. Comparison of Model Accuracy in Dataset 1

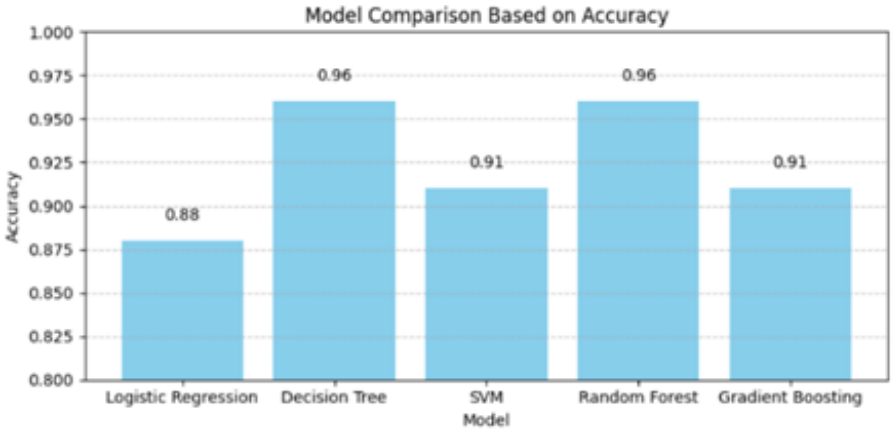


Figure 13. Comparison of Model Accuracy in Dataset 2

5.3 Predictive Accuracy and Business Implications

By comparing the performance of different models on these two datasets, it can be found that the accuracy rates of RF are very high in these two datasets. The accuracy rate of the random forest in Dataset 1 is 95% and recall rate, as well as F1 score are also very good. It means that it will not be able to detect only most of the customers but also those who may be churned. The accuracy rate of RF is the highest in Dataset 2 with a value of 96%. Its accuracy, recall and F1 scores are also quite stable, indicating that it is suitable for predicting whether customers will be churned. These prediction results are very useful for businesses. They can act earlier based on the results made by the prediction. This can increase customer satisfaction and decrease customer demand. In the meantime, it is also capable of decreasing the cost of acquiring new customers and maximizing the profits of companies. Nevertheless, it is also able to make more effective utilization of marketing resources, concentrate on customer in danger of churning, attain accurate marketing and better efficiency.

Other than that, it is also quite important to know the strengths and weaknesses of various models. As an example, there are models whose accuracy rate is higher and yet there is a low recall rate. This can lead to churned customers. To make sure that there is a stable and reliable model, in selecting a model, the F1 score, and the rate of recall should be put into consideration besides the accuracy rate. Overall, an accurate customer churn prediction model could assist the businesses not only in the maintenance of their customer but also in ensuring its growth. It is a relevant instrument of businesses to accomplish management and data-oriented decision-making on their customers.

5.4 Contributions and Comparison with Past Studies

The focus of this study is to develop an effective prediction model for customer churn in the e-Commerce field by using ML technology. The main contributions of this study include a comprehensive comparison of five classification models which are LR, DT, SVM, RF, and GB. To make the results accurate, we will use two different datasets. Overall, Random Forest demonstrates strong and consistent performance across the evaluated metrics, achieving the highest accuracy and F1-score in Dataset 1, while achieving the highest F1-score and precision in Dataset 2. This paper carries out experimental validation that ensemble algorithms like random forest can be robust in various conditions of data, and this method can be used to make customer churn prediction.

Besides, this paper's findings indicate that the RF model predicts customer attrition in two distinct e-Commerce datasets with high accuracy and stability. The RF accuracy rate is 95% in Dataset 1 and 96% in Dataset 2. This shows that the model not only works well but also consistently performs well across a range of user behaviour and interaction data types. This study used two datasets, while earlier paper only used one dataset. For instance, [16] achieved an accurate rate of 96% on the Kaggle dataset using RF, while [18] reached 96.81%. Recent studies have also demonstrated the applicability of machine learning models to e-Commerce customer churn prediction. Hang [24] highlighted the importance of developing robust churn prediction models that can address changing customer behaviour and class imbalance, while Santoso and Sari [25] found that RF provided the most stable performance among the models evaluated for e-Commerce customer churn prediction. Therefore, the consistent RF performance across the two datasets in this study provides additional evidence of its robustness across different e-Commerce data environments. Furthermore, our paper findings further demonstrate the usefulness and effectiveness of random forests, although the paper of [15] and [19] show that XGBoost and ANN are their most successful models. Its performance is even better, especially when compared with models such as LR, GB and SVM. The findings are consistent with recent studies showing that RF can provide reliable performance for e-Commerce customer churn prediction [24], [25]. It is worth noting that RF demonstrated strong reliability in predicting customer churn in e-Commerce by consistently performing on the two datasets used in this paper. To assist e-Commerce platforms in proactively identifying and retaining potential lost customers, this paper not only provides a proven and effective model but also conducts a comprehensive comparison in different data environments. Table 9 shows the result of comparison with past study.

5.5 Model Limitation and Future Work

Although the model used in this paper has a relatively high accuracy rate, there are still some limitations. Among them, the imbalance of customer churn data is the main problem. The number of churn customers is much less than that of non-churn customers, which may affect the performance of the model, especially the recall rate. Dataset 1 comprises 857 non-churn customers (85.7%) and 143 churn customers (14.3%), while Dataset 2 comprises 1,414 non-churn customers (83.72%) and 275 churn customers (16.28%). Therefore, some models tend to predict that customers will not churn, ignoring the actual churn customers. However, simple models such as LR and SVM have poor performance in identifying churn customers, are prone to missed judgments, and have a relatively high false negative rate. Nevertheless, the ensemble model performs relatively well in this study.

To address this limitation, future work should explore advanced techniques for handling imbalanced data. Methods such as SMOTE or cost-sensitive learning can enhance the model’s ability to detect minority class churners, improving recall and overall prediction performance. Additionally, incorporating more diverse and up-to-date customer datasets would help build models that are more generalizable and robust across different e-Commerce environments.

Future research could also investigate hybrid models that combine the strengths of multiple algorithms or explore transfer learning with additional datasets to improve predictive accuracy and stability. Furthermore, applying interpretability techniques such as SHAP or LIME across multiple datasets can help identify consistent churn indicators, providing actionable insights for business decision-making. By following these directions, the predictive framework can evolve into a more reliable, scalable, and business-friendly tool for managing customer churn in e-Commerce.

Table 9. Result Comparison with Past Study

Author	Dataset	ML Model	Accuracy (%)	Best Model
[13]	The dataset is form Alibaba Cloud Tianchi platform	Logistic Regression SVM	LR = 90.66 SVM =91.56	Logistic Regression
[17]	E-Commerce dataset source from Kaggle	Logistic Regression Decision Tree Random forest AdaBoost	LR=88 DT=87 RF=96 AdaBoost=89	Radom Forest
[19]	E-Commerce dataset source from Kaggle	Logistic Regression Decision Tree Random Forest XGBoost LightGBM	LR=78 DT=92 RF=86 XGBoost=96 LightGBM=93	XGBoost
[20]	Dataset from popular e-Commerce platform.	Logistic Regression Decision Tree Random Forest SVM XGBoost AdaBooost	LR=77.35 DT=93.47 RF=96.81 SVM=88.07 XGBoost=96.27 AdaBoost=85.20	Random Forest
[21]	E-Commerce dataset source from Kaggle	Decision Tree Random Forests Naïve bayes Logistic regression SVM K-nearest neighbours Artificial neural network	DT=89.81 RF=91.21 NB=75.04 LR=85.59 SVM=83.30 KNN=84.18 ANN=92.09	ANN
This study	2 e-Commerce datasets that capture user behaviours and interaction activities on the platform.	Random Forest Decision Tree SVM Gradient Boosting Logistic Regression	Dataset 1: RF=95 GB=94 SVM=92 DT=90 LR=86 Dataset 2: RF=96 DT=96 SVM=91 GB=91 LR=88	Dataset 1 RF=95 Dataset 2 RF=96

6. CONCLUSION

This paper explores how to use ML methods to predict customer churn in the e-Commerce industry. The goal is to identify potential customers who may leave so that the businesses can take proactive measures to retain them. Through the evaluation of models such as RF, DT, LR, SVM and GB, the results show that the RF model performs best in terms of their accuracy, precision, recall rate and F1 score.

However, this paper also has certain limitations, such as the problem of data imbalance. Due to the number of churn customers is significantly less than that of non-churn customers, the recall ability of models such as LR and SVM is affected. Although the overall accuracy performance of each model is relatively good, there is still room for improvement in its comprehensive performance and generalization ability. To solve these problems, it is recommended to adopt techniques such as SMOTE to mitigate the impact caused by data imbalance. Elsewhere, SMOTE can be further explored and future research could also examine cost-sensitive learning and ensemble

resampling techniques that adaptively adjust for skewed datasets, ensuring that minority churn cases are detected more reliably.

Overall, this paper proves that ML is effective in predicting customer churn. The research results provide practical methods for enterprises to help reduce customer churn. Meanwhile, it also lays a foundation for future research and practical application. As long as the data quality is constantly improved and the stability of the model is enhanced, these predictive tools can better help enterprises retain customers and achieve long-term development. Another important direction is the adoption of interpretability techniques such as SHAP and LIME, which can provide stakeholders with insights into the decision-making process of ML models, particularly Random Forest. These tools can highlight the most influential features driving churn prediction, allowing businesses to design more targeted retention strategies.

For reproducibility and transparency, future work should ensure that preprocessing steps, feature engineering, and model training parameters are explicitly documented, as these factors can significantly affect results. Expanding the analysis to hybrid models that combine the strengths of different algorithms or applying the framework to additional and more diverse datasets, could further improve robustness and generalizability. Extensions of this study might also investigate temporal data, customer sentiment, or cross-platform behaviours to gain deeper insights into churn dynamics. By integrating these approaches, customer churn prediction research can evolve into a more interpretable, scalable, and practically useful tool for e-Commerce businesses.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their valuable comments.

FUNDING

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Tan Yi Qi: Data Preparation, Modelling, Validation, Writing – Original Draft Preparation;
Lew Sook Ling: Conceptualization, Supervision, Review – Editing;
Ram Avtar: Conceptualization, Review – Editing.

CONFLICT OF INTERESTS

No conflicts of interest were disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author on request.

REFERENCES

[1] Y. Feng, Y. Yin, D. Wang, J. Ignatius, T. C. E. Cheng, M. Marra, and Y. Guo, "Enhancing e-Commerce customer churn management with a profit- and AUC-focused prescriptive analytics approach," Journal of Business Research, vol. 184, Art. no. 114872, 2024, doi: 10.1016/j.jbusres.2024.114872.

[2] D. Raut, A. Nimje, D. Bhagwatkar, N. Rakesh, R. Sharma, and M. Gulhane, "Predicting customer churn rate using machine learning," in 2025 10th International Conference on Signal Processing and Communication (ICSC), Feb. 2025, pp. 529–534. doi: 10.1109/ICSC64553.2025.10968946.

[3] N. Danaan, A. Deme, G. Bibu, M. A. Lawal, and I. Z. Yakubu, "An integrated machine learning model for e-Commerce churn prediction," *International Journal of Innovative Science and Research Technology,* vol. 10, no. 6, pp. 2762–2778, Jul. 2025. doi: 10.38124/ijisrt/25jun1816.

[4] S. J. Shobana, Ch. Gangadhar, R. K. Arora, P. N. Renjith, J. Bamini, and Y. D. Chincholkar, "E-Commerce customer churn prevention using machine learning-based business intelligence strategy," *Measurement: Sensors*, vol. 27, Art. No. 100728, Jun. 2023. doi: 10.1016/j.measen.2023.100728.

[5] I. Jahan and T. F. Sanam, "An improved machine learning-based customer churn prediction for insight and recommendation in e-Commerce," in 2022 25th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh: IEEE, Dec. 2022, pp. 1–6. doi: 10.1109/ICCIT57492.2022.10054771.

[6] R. Alexander, A. Maria Nancy, E. Aswini, and P. S. Sarao, "Comparative performance analysis using machine learning for churn prediction in e-Commerce," in 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India: IEEE, Dec. 2023, pp. 537–542. doi: 10.1109/ICACRS58579.2023.10404319.

[7] S. Upadhyay and M. Rajalakshmi, "Customer churn prediction using machine learning," in 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India: IEEE, Jul. 2023, pp. 1–7. doi: 10.1109/ICCCNT56998.2023.10307748.

[8] J. Sharma and S. Neema, "CusCP: AI-driven system for predictive modeling of customer churn in e-Commerce using machine learning," in 2023 Seventh International Conference on Image Information Processing (ICIIP), Solan, India: IEEE, Nov. 2023, pp. 924–929. doi: 10.1109/ICIIP61524.2023.10537663.

[9] IBM, "What is machine learning (ML)?," IBM Think, Sep. 22, 2021. [Online]. Available: https://www.ibm.com/think/topics/machine-learning

[10] M. A. Al Rahib, N. Saha, R. Mia, and A. Sattar, "Customer data prediction and analysis in e-Commerce using machine learning," *Bulletin of Electrical Engineering and Informatics,* vol. 13, no. 4, pp. 2624–2633, Aug. 2024, doi: 10.11591/eei.v13i4.6420.

[11] IBM, "What Is a Machine Learning Algorithm?," *IBM Think*, Nov. 3, 2023. [Online]. Available: https://www.ibm.com/think/topics/machine-learning-algorithms

[12] M. Afzal, S. Rahman, D. Singh, and A. Imran, "Cross-sector application of machine learning in telecommunications: enhancing customer retention through comparative analysis of ensemble methods," IEEE Access, vol. 12, pp. 115256–115267, 2024, doi: 10.1109/ACCESS.2024.3445281.

[13] X. Xiahou and Y. Harada, "B2C e-Commerce customer churn prediction based on K-Means and SVM," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 2, pp. 458–475, Apr. 2022, doi: 10.3390/jtaer17020024.

[14] A. R. Lubis, S. Prayudani, J. Polmed, O. Nugroho, Y. Y. Lase, and M. Lubis, "Comparison of model in predicting customer churn based on users' habits on e-Commerce," in 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia: IEEE, Dec. 2022, pp. 300–305. doi: 10.1109/ISRITI56927.2022.10052834.

[15] P. Aggarwal and V. Vijayakumar, "Customer churn prediction in the telecom sector," in 2024 3rd International Conference on Artificial Intelligence for Internet of Things (AIIoT), Vellore, India: IEEE, May 2024, pp. 1–6. doi: 10.1109/AIIoT58432.2024.10574660.

[16] S. Kim and H. Lee, "Customer churn prediction in influencer commerce: an application of decision trees," Procedia Computer Science, vol. 199, pp. 1332–1339, Jan. 2022, doi: 10.1016/j.procs.2022.01.169.

[17] S. Brinthakumari, K. Patil, K. Pawar, and S. Jogale, "Customer churn prediction using machine learning," *International Journal of Scientific Research in Engineering and Management (IJSREM)*, vol. 8, no. 4, pp. 1–5, Apr. 2024. doi: 10.55041/IJSREM29907.

[18] G. Chaubey, P. R. Gavhane, D. Bisen, and S. K. Arjaria, "Customer purchasing behavior prediction using machine learning classification techniques," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 12, pp. 16133–16157, Dec. 2023, doi: 10.1007/s12652-022-03837-6.

[19] J. Li, "Customer churn prediction using machine learning: a case study of e-Commerce data," *International Journal of Computer Applications,* vol. 186, no. 48, pp. 22–25, Nov. 2024, doi: 10.5120/ijca2024924140.

[20] T. M. Ghomeeda, "Utilizing machine learning algorithms for the comprehensive prediction and analytical study of customer churn in electronic commerce," in *ICSE 2024*, Libya: BWU, 2024. [Online]. Available: http://bwu.edu.ly/icse2024.

[21] I. Boukrouh and A. Azmani, "Explainable machine learning models applied to predicting customer churn for e-Commerce," *IAES International Journal of Artificial Intelligence,* vol. 14, no. 1, pp. 286–297, Feb. 2025, doi: 10.11591/ijai.v14.i1.pp286-297.

[22] "Digital Marketing | E-Commerce | Customer Behavior," *Kaggle*, [Online]. Available: https://www.kaggle.com/datasets/ermismbatuhan/digital-marketing-ecommerce-customer-behavior. Accessed: Jul. 5, 2025.

[23] "Ecommerce Customer Churn Analysis and Prediction," *Kaggle*, [Online]. Available: https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction. Accessed: Jul. 5, 2025.

[24] J. Hang, "Customer churn prediction model for E-Commerce using an enhanced adaptive boosting algorithm," *Discover Applied Sciences*, vol. 8, Art. no. 747, 2026, doi: 10.1007/s42452-026-08745-4.

[25] R. B. Santoso and B. W. Sari, "Analisis simulasi prediksi customer churn e-Commerce menggunakan algoritma Random Forest berbasis data sintetis," *JMBI UNSRAT (Jurnal Ilmiah Manajemen Bisnis dan Inovasi Universitas Sam Ratulangi)*, vol. 13, no. 1, pp. 317–331, 2026, doi: 10.35794/jmbi.v13i1.67545.

BIOGRAPHIES OF AUTHORS

	**Tan Yi Qi** received the bachelor's degree in information technology, specializing in business intelligence and analytics, from Multimedia University. Her research interests include customer churn in e-Commerce, machine learning methods, and predictive analytics. She can be contacted at email: 1211202138@student.mmu.edu.my.
	**Lew Sook Ling** is an Associate Professor at the Faculty of Information Science and Technology, Multimedia University (MMU), Malaysia. She has been with MMU since 2001 and received her Ph.D. in 2013. She is also a Senior Member of IEEE. Her current research interests include educational technology, business analytics, image processing, and machine learning. She actively contributes to academic development through publications and collaborative research in emerging technologies. She can be contacted at email: sllew@mmu.edu.my.
	**Ram Avtar** received the Ph.D. degree in civil engineering from The University of Tokyo, Japan. He is currently an Associate Professor with the Faculty of Environmental Earth Science, Hokkaido University, Japan, and the Director of the Global Land Program (GLP), Japan Nodal Office. He is actively contributing to study land systems and co-design solutions for global sustainability as a part of GLP Program. He was with the Institute for the United Nations University (UNU-IAS), as a Research Fellow for four years (2012–2016). He has developed methods for mapping natural resources using multi-sensor remote sensing techniques and scenario analysis for sustainable management of these resources. He can be contacted at email: ram@ees.hokudai.ac.jp.