Measurement: Sensors 27 (2023) 100728

Contents lists available at ScienceDirect

Measurement: Sensors

journal homepage: www.sciencedirect.com/journal/measurement-sensors

E-commerce customer churn prevention using machine learning-based business intelligence strategy

Shobana J [a,*], Ch. Gangadhar [b], Rakesh Kumar Arora [c], P.N. Renjith [d], J. Bamini [e], Yugendra devidas Chincholkar [f]

[a] Department of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamilnadu, India
[b] Department of Electronics & Communication Engineering, Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, 520007, India
[c] Department of Information Technology, Krishna Engineering College, Mohan Nagar, Ghaziabad, UP, India
[d] Computer Science and Engineering Department, Vellore Institute of Technology, Chennai, India
[e] Sri Ramakrishna College of Arts & Science, Coimbatore, Tamil Nadu, India
[f] Department of Electronic & Telecommunication, Sinhgad College of Engineering, Pune, India

A R T I C L E I N F O

Keywords:
Index terms— E-commerce customer churn
Hybrid algorithm
Personalized retention
Support vector machine

A B S T R A C T

Businesses in the E-Commerce sector, especially those in the business-to-consumer segment, are engaged in fierce competition for survival, trying to gain access to their rivals' client bases while keeping current customers from defecting. The cost of acquiring new customers is rising as more competitors join the market with significant upfront expenditures and cutting-edge penetration strategies, making client retention essential for these organizations. The best course of action in this circumstance is to detect prospective churning customers and prevent churn with temporary retention measures. It's also essential to understand why the customer decided to go away to apply customized win-back strategies. Each customer's information, including searches made, purchases made, frequency of purchases, reviews left, feedback is given, and other data, is kept on file by the e-commerce company. Machine learning and data mining may be aided by examining this enormous quantity of data, analysing customer behaviour, and seeing potential attrition opportunities. The support vector machine is a popular supervised learning method in machine learning applications. Predictive analysis uses the hybrid classification approach to address the regression and classification issues. The process for forecasting E-Commerce customer attrition based on support vector machines is presented in this paper, along with a hybrid recommendation strategy for targeted retention initiatives. You may prevent future customer churn by suggesting reasonable offers or services. The empirical findings demonstrate a considerable increase in the coverage ratio, hit ratio, lift degree, precision rate, and other metrics using the integrated forecasting model. To effectively identify separate groups of lost customers and create a customer churn retention strategy, categorize the various lost customer types using the RFM principle.

1. Introduction

Internet technology has tremendously impacted customer lives because of the growth and process of network technology and social information. With the advent of "Internet +" e-commerce, traditional businesses must use the Internet platform to open up new markets and profit growth points, and online companies drive the rapid growth of the retail market in the network [1]. The development of the internet retail business has increased industry competition. Because the customer churn rate in the e-commerce industry is so high, business owners must examine ways to reduce customer churn in online purchasing. Because the customer's behaviour is predictable, it is possible to predict the customer's future trading inclinations using the relevant data acquired and the necessary analysis [2]. To limit the number of lost clients, business owners can detect the consumer who has a losing tendency and perform the required pre-control work. Customer attrition prediction for online buying has become a significant focus of e-commerce company research in recent years [3]. Client churn, also known as attrition, occurs when an existing customer ends their association with a company by no longer using its services or purchasing its products. Customers in an

* Corresponding author.
E-mail addresses: shobanaj1@srmist.edu.in (S. J), gangadharch1111@gmail.com (Ch. Gangadhar), dr.rakeshkarora@gmail.com (R.K. Arora), renjith.pn@vit.ac.in (P.N. Renjith), baminikar@gmail.com (J. Bamini), ydc2002@rediffmail.com (Y. Chincholkar).

https://doi.org/10.1016/j.measen.2023.100728
Received 16 December 2022; Received in revised form 14 February 2023; Accepted 2 March 2023
Available online 8 March 2023

E-Commerce situation can be in one of four states: new, active, inactive, or churned [4]. The industry's success is entirely based on its capacity to keep customers engaged for an extended period. Because of the industry's fierce rivalry, obtaining a new consumer is relatively expensive. In the event of a new customer, a firm's financial breakeven (return on investment) is usually reached only after the customer completes a few transactions over time [5]. Active consumers are the backbone of every E-Commerce business, and businesses should pay special attention to those who may become inactive and churn later. Statistical techniques and machine learning strategies [6] are examples of methodologies that can be used to anticipate prospective client turnover. Customer data collected by B2C E-Commerce companies are a great source of information because it allows for analyzing different consumers' purchasing patterns. The likelihood of a client churn is expected via churn prediction. It reduces the cost of acquiring new customers while also assisting in customer retention. It takes more marketing time and money to develop a new client than to keep an existing consumer [7]. Customers who are hesitant to make a purchase or are prepared to switch shopping sites due to financial concerns can be persuaded and clutched. They can expect standards and variety in product offerings. Customers leaving for essential and unavoidable reasons are free to do so. Though we invest in involuntary churners, the result is solid. Target marketing assists in reaching and connecting with customers [8]. In a contract scenario, a customer relationship indicates that the firm and the client work together [9]. Both parties' rights and obligations are clearly stated in the contract. The client must complete the necessary responsibility following the contract to enjoy the appropriate freedom after signing an agreement with the firm [10]. Extending discounts, changing products to consumers' preferences, and sending out trigger emails are all ways to discourage voluntary churners. Focusing solely on voluntary churners will reduce the cost of providing advantages to new and existing churned consumers [11]. Predicting client attrition with unstructured data produces poor outcomes and

Creates a problem of class imbalance. To prevent these issues, there must be a difference significant in the percentage of churn in historical data and non-churn in historical data. Customer data should be preprocessed, and feature selection should be used to forecast the essential properties. This saves time and money when it comes to forecasting. Different classification and predictive models can be used to predict customer attrition. Efficient algorithms focused on soft voting on accuracy are exposed, which chooses the ensemble model as the best to follow in future works [12].

1.1. Related works

Researchers have taken a variety of approaches to predict churn. A conceptual model to investigate the effect of consumer loyalty and relationship quality on the client maintenance process [13] was one of the principal concentrates around here. The age of programmed maintenance methodologies, as depicted by Bolton et al. furthermore Hadden et al., is a sort of suggestion administration. There are various examinations in the field of recommender frameworks, including one that analysed the abilities of three information mining models - neural organizations, relapse, and relapse trees in gauging client churn [14]. A combination of SMC models and naïve Bayesian forecasts the performance of e-commerce website customers. The new model, the Naive Bayesian method, outperforms the SMC model. The shopper movement level of an online business site utilizing the non-authoritative relationship and the SMC model. As indicated by the information, the higher a client's degree of movement, the lesser the possibility of client churn Wu Hong utilizes the SMC model to figure forthcoming client esteem and incorporate it into client property attributes [15]. Credit cards, Retail banking, mobile internet gambling, social gaming, and the email marketing are some of the other areas of expertise. Scholars have developed unique churn prediction models and studies in a variety of industries, such as telecom, where a lot of work has been done, such as. In previous work on B2C E-Commerce customer churn detection, customer churn prediction was obtained using a logistic regression approach [16]. The customer profile is one of the most obvious input criteria for recommendation generation, and previous recommender systems relied heavily on it, as well as the user's behavioural features. Further studies began to look into things like trust factor historical transaction patterns, societal inputs from people in comparable demographic groups and/or with similar past transaction qualities, contextual information, and so on [17]. It is called a collaborative method when social inputs are taken into account when making recommendations. This method is usually used in conjunction with a demography-based method [18]. The similarity of client profiles, which is computed via computational methodologies, is a significant parameter to consider in such models. Customer retention and loyalty programme memberships were explored as a result of service experiences. Neslin and colleagues investigated the methodological components that determine the accuracy of customer churn prediction models in a descriptive study. Jamal et al. created a model to examine the relationship between influencing elements such as failure recovery, customer attrition, and customer service experience [19]. The study of customer behaviour in non-contractual relationships focused mostly on buying behaviour. The SMC model for predicting consumer transactions was proposed by Schmittlien Morrison and Colombo in 1987 [20]. The model used mathematical calculations to determine the customer's activity and then used that information to predict the customer's behaviour in the future [21].

2. Methodology

The use of labelled training datasets in supervised learning is a type of machine learning. These datasets are used to train or support vector machine algorithms, allowing them to reliably identify data and predict outcomes. Using labelled inputs and outputs, the model may be tested for accuracy and learn over time. Artificial intelligence and machine learning, which has pervaded the BI sector and is transforming the way businesses think about their data, is at the forefront of this automation. Machine learning, like any other technical frontier, may be a risky topic for enterprises. In order to deal with client turnover in B2C E− Commerce, the technique proposed in this study adopts a two-pronged approach. The first component employs support vector machine technology to anticipate churn occurrences. The second portion looks at how to use a hybrid strategy that blends content-based, collaborative, demographic, and knowledge-based methodologies to produce customised retention approaches. Business intelligence is a system that collects, analyses, and presents massive amounts of data and information. When used correctly business intelligence solutions and products save you time and energy while providing you with crucial information about your Business and clients. Business intelligence software can parse the bulk of your raw data understand it and present you with the information that matters. Organizations can use business knowledge to make increasingly knowledgeable sales, better decisions in marketing and other areas of the Business. E-commerce enterprises can employ business intelligence to make data-driven decisions. It is possible to settle on choices about future firm development by taking a long term at a drawn-out perspective available, market, contest, and different realities. Business knowledge can be utilised by web-based business associations to distinguish explicit deals patterns dependent on their clients' preferences, internet shopping encounters, habits, buying conduct, and responses to advancements and patterns, all of which impact deals and assist them with augmenting income. Any's organization will likely decrease the waste while expanding income. Business knowledge can help E-trade associations distinguish flaws for quality concerns including client steady loss, lost deals usefulness because of call focus discontent, poor technique because of ineffectively investigated statistical surveying, and a scope of different hardships that request quick efficient answers. Therefore, information might be prepared through an assortment of modules to recognize waste improvement openings and set up plans to address these worries. Business Intelligence helps E-trade organizations in refining a different stock and enhancing supply sums by evaluating considerable recorded information, for example, client profiles and buy purchase habit. It also decreases the chance of out-of-stock situations by analysing safety stock data, sales, and data inventories to generate accurate estimates. Business intelligence can predict overstock situations before they become a severe problem by merging sales, replenishment, and forecasting data.

2.1. *The concept of customer churn*

Customer churn refers to a situation in which a customer's contribution to the company's earnings is decreasing. Client churn is the context of contractual relationships and non-contractual relationships are divided into two groups based on whether the company and the customer engaged into a contract.

With regards to a legally agreement for understanding the customer churn. In an agreement situation, a client relationship shows that the firm and the customer cooperate to bring down the two parties' misfortunes in the exchange cycle and to tie the agreement by restricting the agreement. The two parties' privileges and commitments are obviously expressed in the agreement and the relating just subsequent to marking an agreement with the contract the customer should finish the important duty as per the agreement. The exchange between the client and the firm is limited by the understanding in the contract relationship, the customer should pay a more prominent expense of move, and the client's exchanging conduct is sensibly predictable. Telephone communication, protection organizations, etc are instances of agreement-based organizations. With regards to non-contractual relationship organization, client churn is an issue. The relationship between the firm and the client begins with the customer's initial transaction in the case of non-contractual situations. Customers and businesses do not need to sign a contract; customers can easily enter and exit the Business, and the trade and loss behaviour of customers is unpredictable. Customers may buy things for a long period after their first purchase or never trade again. The customer turnover rate is higher in the setting of non-contractual relationships because the firm is less bound to the customer, combined with the minimal amount of client transfer. Into non-contractual partnerships, measuring customer turnover is more challenging since the reasons for lost consumers are more ambiguous, and there is no strong division between lost and non-lost clients. E-commerce businesses must establish a benchmark for customer turnover based on the characteristics of their products in order to detect the trend of lost clients in a timely manner and adopt effective retention strategies. The Fig. 1 provides the overview customer churn management.

Customers who shop online are characteristic of customer churn in non-contractual relationships. When the length of the transaction is considered, the failure of an online shopping client is split into two categories: permanent loss and interruptions.

(1) **Loss that occurs on a regular basis**: The term "intermittent loss" refers to clients who did not purchase the company's goods or services over a set length of time, with the main feature being a reduction in trade frequency. Customers who fall into this category may not be able to purchase the company's product within the specified time limit. The customer is not completely lost; he or she may still acquire the company's service. Customers in this situation may be unable to buy the firm's manufactured goods within the specific time frame. However, this does not imply that the customer is absolutely helpless. After the time limit has expired, the customer may purchase the company's product or service.

(2) **Completely loss**: The term "permanently lost" refers to a customer's decision not to purchase services or the company's products in the future. E-commerce businesses will not delete a client's account; even if the customer hasn't been interested. The consumer can log in using the registered account for a long period the Business will not be able to tell whether the customer's account is permanently gone. Permanent loss refers to a client's complete loss, which can occur for a variety of reasons, including a change in the customer's purchasing habits, as well as a change in the growth stage at which the product is no longer required.

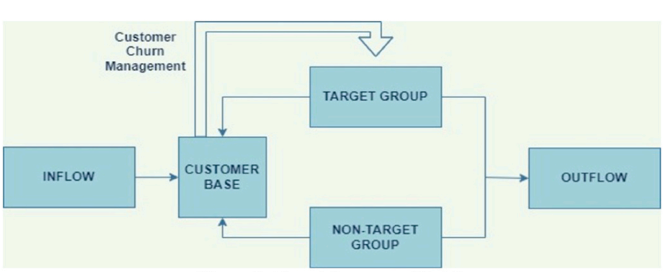

Fig. 1. Churn management model.

2.2. *Detection of customer churn*

Statistical models and machine learning algorithms are examples of predictive analytics techniques can be used to populate risk scores for each client, allowing for proactive churn detection. Usage data, customer reviews, customer demography, net promoter score, purchasing patterns, and other input characteristics are all used in this process. Since the date on which a customer registers with a B2C E-Commerce firm, all of this information has been available to them. So, the next step is to finalise a framework for processing this data. One possibility is to utilise statistical models based on regression techniques such as logistic regression to anticipate prospective churners. Another method is to use machine learning technologies such as the support vector machine. The Support Vector Machine is a supervised learning technique that may be used to solve both regression and classification issues in machine learning. SVM, which splits a set of entities into distinct class memberships, is supported by the decision planes idea. SVM is prepared on a bunch of information in which every component is characterized into one of two classes. SVM makes a model that can relegate each new member to one of the two classes depending on the data obtained. To do this, an algorithmically developed ideal hyperplane in the element space is made, filling in as a choice limit for the arrangement issue. SVM can be considered a non-probabilistic paired direct classifier in this situation shown in Fig. 2.

For a given data collection (X1, X1), (X2, X2), (X3, X3), (Xm, Xm), Yi = −1 for inputs in class Xi = 0 and Yi = 1 for inputs Xi in class 1, where Yi = 1 for inputs Xi in class 1.

The decision boundary can be expressed mathematically as a vector representing a two-dimensional line and two-dimensional vectors, and b

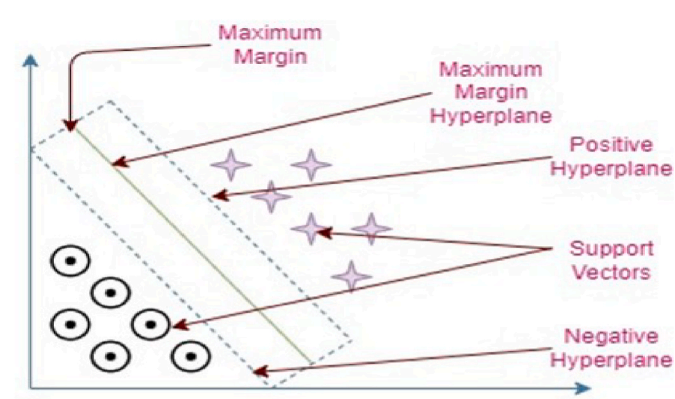

Fig. 2. SVM maximum margin hyperplane.

represents the bias as a constant. In this way, the negative support vector n can be defined as the input vector from class 0, and the positive support vector p may be defined as the input vector from class 1.

Support Vector Machines use a technique known as the kernel trick to achieve nonlinear classification. In the high-dimensional feature space input components are mapped and the kernel trick is used to perform a classification task. The method will be similar to that used in the linear case, with the exception that the nonlinear kernel function will be used to replace every dot product. The decision boundary (also known as the maximum-margin hyper plane) can now be fitted using algorithm converts the feature space based on higher-dimensional Format. The hyperplane of the revised feature space may become nonlinear when remapped to the original input space in Fig. 2. Mercer's condition can be seen as a kernel function for any mathematical function that can verify as a general rule.

On the other hand, the most often used kernel functions are based on Euclidean inner products or Euclidean distance. The type of class boundaries and the data structure employed are essential elements that influence kernel selection. SVM is regarded as the most reliable classification technique in various difficult real-time circumstances. Handwritten character recognition image/face detection, hypertext categorization and fraud detection are a few examples. In these cases, the main benefit of SVM is that it avoids the overfitting problem by using an appropriate training phase to generate the classification model.

In the case of Business to consumer for E-Commerce churn detection, a binary classification of the entire client base using a support vector machine is used to identify prospective churners in Fig. 3. Each of the client characteristics n-dimensional vector space is represented with each of the consumer qualities is represented by a vector dimension in the feature space. Historic churn data, customer characteristics, and transaction traits are used to train the SVM Model showing in Fig. 3. A periodic churn test is also run for each customer database member to detect probable attrition scenarios.

2.3. Retention action

In the B2C E-Commerce industry a hybrid method is the most dependable way for generating a list of best tailored retention alternatives. To achieve the best recommendation result, it is advised to combine, collaborative, content-based, knowledge-based, and demographic techniques shown in Fig. 4.

Hybrid classification algorithm:

The overview of the proposed algorithm

Step 1. Determination of the various training data set.

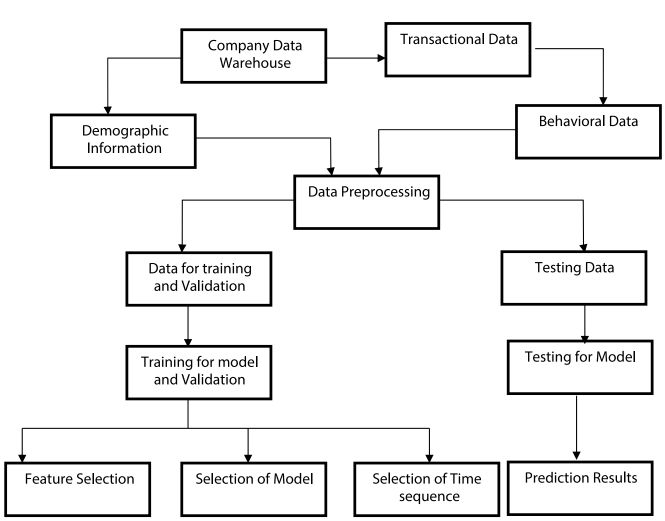

Fig. 3. Customer Churn Detection Framework based on SVM.

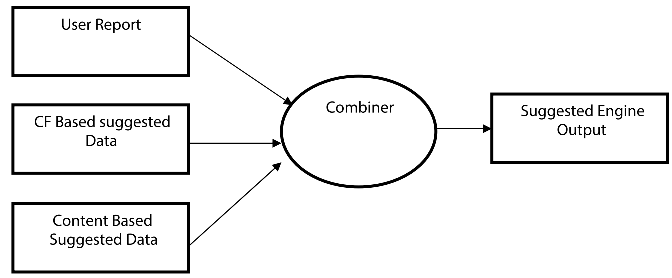

Fig. 4. Retention Actions based Hybrid Recommendation Strategy.

Step 2. Determination of the various test data set.

Step 3. Determination of the calculated accuracy.

Step 4. Selection of the optimal value of cost and gamma.

Step 5. Implementation of the SVM train step for every data point.

Step 6. Implementation of the SVM to classification of the testing data points.

Step 7. Returning of the accuracy value.

The customer's personal information and previous transactional qualities serve as the input data for creating content-based recommendations. The customer's demographic data and their accumulated expertise over time serve as the inputs for knowledge- and demographic-based recommendations. Collaborative suggestions are created using the use habits of similar users. To make similarity-based recommendations, similarity functions are utilised to calculate the perceived value of the recommendations to the target consumer. Recommendations generated in this way often have a high level of personalization since they are unique for each customer and are influenced by their prior activity patterns and individual profiles.

Customer While "j" is a client belongs to the set of J's is loyal customers, whose set of customers be J with more than 90% loyalty, i.e., a churn probability of less than 10%, I belong to the predicted set of potential attrition candidates. In contrast to conventional recommender systems, which always construct the neighbourhood with customers who are exactly like the client in question, this approach to neighbourhood creation uses consumers who are different from the client. The following are the main factors or assumptions to take into account while creating a customised retention strategy for a likely churn scenario: You may prevent future client turnover by proposing appropriate offers and services. The value of an offer or service may vary based on the consumer, and customers with similar tastes will behave similarly.

2.4. Processing of information

The data obtained by online crawling technology is imperfect and nonlinear, and the information gathered cannot be directly used to the Research. To satisfy the actual research needs, it must be standardised to deal with the original data. The sole constant in the development of online operations is the customer's method-assigned ID. To efficiently distinguish various consumers, an ID corresponding to the unique number is employed. Because transaction time is a single point in time, it's critical to transform it to a continuous variable and standardise it to [0, 1]. Because the prices of the items and the numerical score show considerable differences, data normalisation is required to reduce the data gap on the influence of the result, as well as data normalisation to [0,1].

Considering the characteristics of electronic objects shown in Fig. 5, this study uses client exchange information in one year such as the observation period and client exchange data from one year as the projection period. Customers that don't make a purchase within the anticipated period are categorised as "1″ lost customers, while those who do

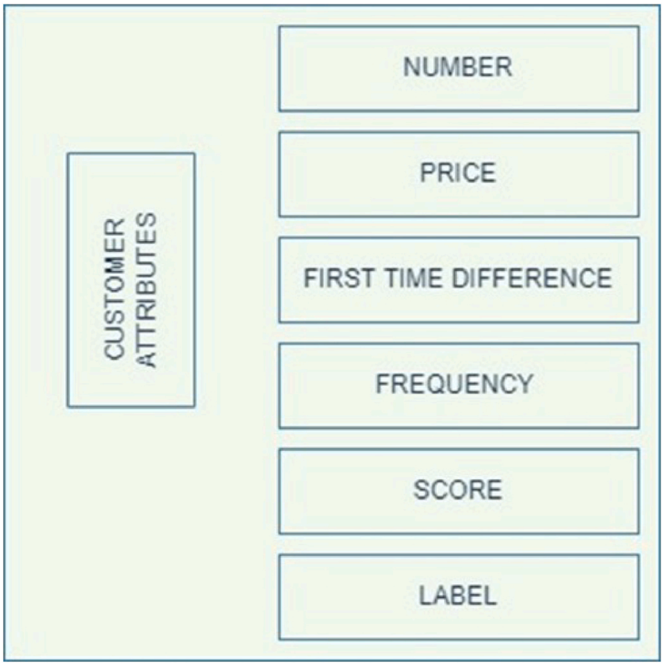

Fig. 5. Attributes for customer.

make a purchase during the anticipated period are categorised as "0″ customers. The customers ascribe contain the customer's amount, the item's price, and a description of the item. The first time the client exchanged as First TimeDiff, the last time the client exchanged as TimeDiff, the recurrence of exchanging, the client's score, and the client's class (Label). Table 2 shows the client's property table.

2.5. Data collection

The difference between the number of lost and non-lost clients is taken into importance while obtaining the training data set. This Research simultaneously employs a systematic sampling technique to extract an equal number of lost and non-loss consumers in order to fully extract the properties of lost customers shown in Fig. 6 the entire number of clients—table customers included—both lost and not lost. According to the present empirical investigation, the training sample and test sample are used in a ratio of 2:1 in terms of number of training sample and test sample.

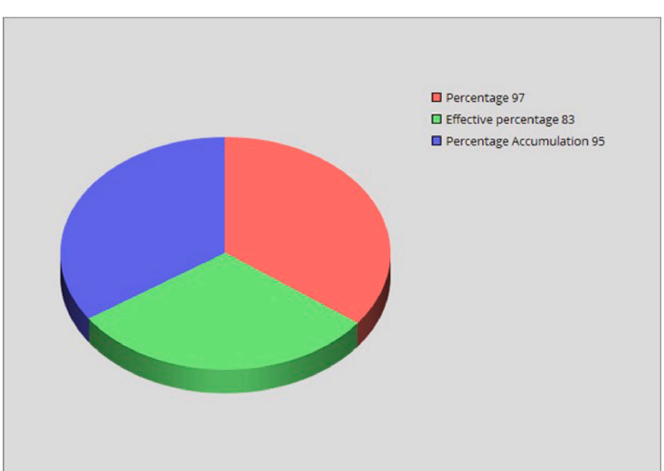

Fig. 6. Analyzing the data from the training sample.

2.6. Analyze the data

Before the model is created, the relationship between customer churn and customer attributes is investigated, as well as the influence of customer attributes on customer turnover. The link between commodities prices and churn among customers Fig. 7.

The horizontal axis shows the price of the goods bought by the customer. The graph shows that much of the time, the price of purchased products falls within the range of 0–0.4. People are more inclined to buy electronics at a cheaper price, whereas those at a higher price are less likely to do so. As commodity prices rose, the percentage of clients who left the company decreased. An association is drawn between the initial transaction a consumer makes and subsequent customer attrition shown in Fig. 8.

For example, the horizontal axis shows how long it took the consumer to begin trading. The relationship between customer attrition and trading behaviour was not apparent at first glance. There's no rhyme or reason to the two being linked. customers who have previously traded on a regular basis are more likely to leave the company in Fig. 9.

The horizontal axis of the customer's trading frequency is depicted in the Fig. 9, with both lost and non-lost clients clustered in the 0 to 0.3 range. As seen in Fig. 10, as customer purchases is increase then the fraction of non-lost consumers increases at the same rate. The link between customer churn and a product's score.

On the horizontal axis, which displays the customer's product score, both lost customers and non-lost customers are mostly grouped in the lower score range. On the other side, when customers' commodity scores increase, the percentage of lost customers is decreasing.

2.7. Model establishment

Model establish the clear linear relationship between customer churn and customer qualities, according to the exploratory Research of customer attributes and customer churn. It can take into account an artificial neural network's ability to cope with classify customer turnover and nonlinear data effectively. When the number of clients in an E-commerce business is reasonably steady over time, it can use the support vector machine model to classify small samples of data. Composite forecasting has been more popular in recent years, with several researchers combining artificial neural networks, genetic algorithms, and other methods. To predict the condition of client churn dependent on the traits of internet shopping client exchange information, use support vector machines and artificial neural organizations for coordinated figuring out how to play for their separate potential benefits seen in Fig. 11.

Client churn between January and February was predicted using the

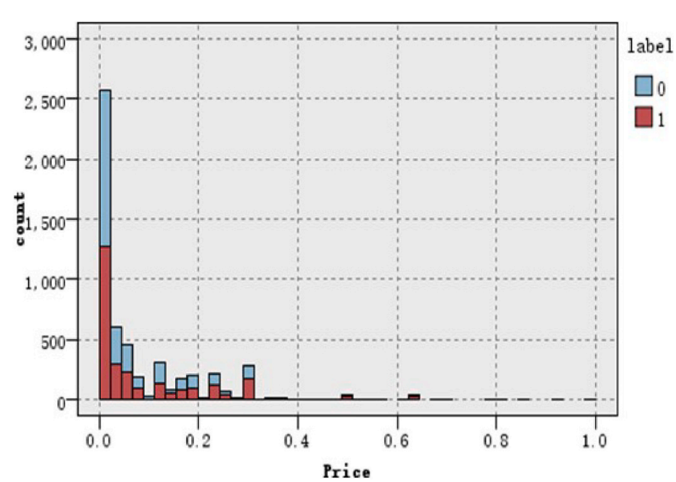

Fig. 7. The connection between pricing and consumer attrition.

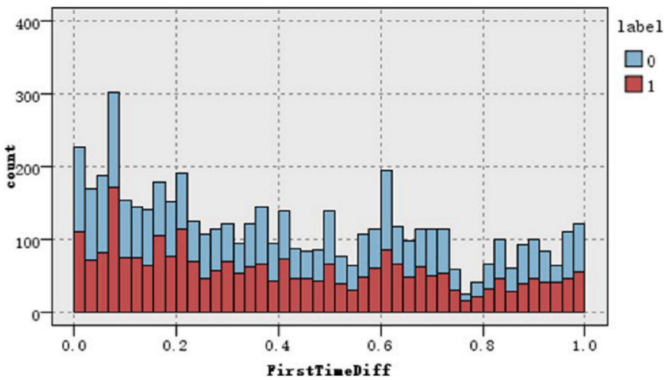

Fig. 8. Customer churn and the relationship between a client's initial trading activity.

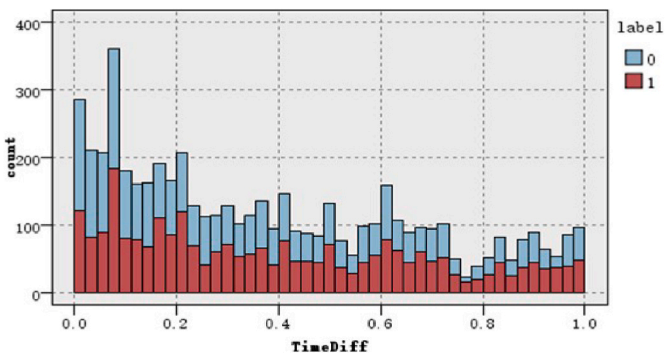

Fig. 9. The connection between trading frequency and client attrition.

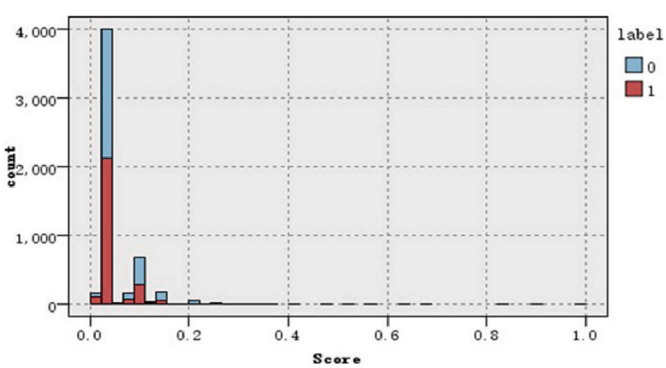

Fig. 10. The relationship between customer churn and product's score.

Results for output field label

Comparing $N-label with label

Correct	1,347	82.64%
Wrong	283	17.36%
Total	1,630	

Coincidence Matrix for $N-label (rows show actuals)

	0	1
0	44	121
1	162	1,303

Fig. 11. The BP artificial neural network's findings.

BP artificial neural organisation model training datasets, with an accuracy rate of 82.64% and an error rate of 17.36% seen in Fig. 12.

2.8. Model assessment

The customer churn customer table is a visual representation of the true characterisation of lost and non-lost customers in real forecasting. A customer churn model may be assessed by choosing the coverage rate, hit rate, lift degree, and accuracy rate are shown in Table 1.

The lift curve is made up of three lines: a baseline at the bottom, an ideal curve at the top, and an actual lift curve in the middle. The baseline is the shortest line. When the lift curve is closer to the ideal curve, the model performs better. When compared to the improvement curves of the separate models, the combined model's lifting curve is shown to be close to the ideal curve. Individual forecasting models do not perform as well as a combined forecasting model shown in Fig. 13 and Fig. 14.

2.9. Performance analysis

For this Research, we use the PEN hypothesis, in which P stands for recent purchase rate, E stands for frequency of exchanges, and M stands for the purchase amount. In the absence of the customer's commitment to the association's value, N stands for monetary. As a result of the adjustments made to the three loads, the client's value increased. Because of the time hub's influence on the client's most recent purchase date, the two components F and M are split into two states. N2 is more costly than the average item, whereas N1 is less expensive than the average item price. In contrast to E2, which happens more often than the average exchange rate, E1 occurs less often. The PEN hypothesis is utilised to estimate the worth of lost consumers in this investigation shown in Table 2.

More than half of all lost clients are E_1N_1 customers, making them the most prevalent sort of client. Even if such clients might be dismissed due to their lower value in a conventional administration strategy, online company firms should not disregard such consumers. Businesses engaged in e-commerce need to address the problem of decreasing the loss of low-value customers. clients who have the E_1N_2 code. Although this kind of consumer trades less often than the average, they do so more frequently than the average, with a volume that is around 25.5% higher. The trading frequency is approximately the same for E_1N_1 customers. 9.8% of E_2N_1 customers in this category have experienced financial loss. As the purchase value stays the same, increasing customer purchase frequency significantly lowers turnover when compared to E_1N_1 consumers. Increasing the quantity of client purchases is harder after a while, when the customer's revenue is essentially steady. The best strategy is thus to increase customer frequency. Consumers of E_2N_2 have the lowest rate of lost customers, regardless of whether they purchase more often or in larger quantities than the norm. Increasing the quantity of things sold and the trading frequency of online shoppers would

Results for output field label

Comparing $S-label with label

Correct	1,261	77.36%
Wrong	369	22.64%
Total	1,630	

Coincidence Matrix for $S-label (rows show actuals)

	0	1
0	63	102
1	267	1,198

Fig. 12. The output of a support vector machine.

Table 1
The customer churn categorization table.

Actual Prediction	Non Lost Customer	Lost customer
Non Lost Customer	f_{11}	f_{12}
Lost customer	f_{21}	f_{22}

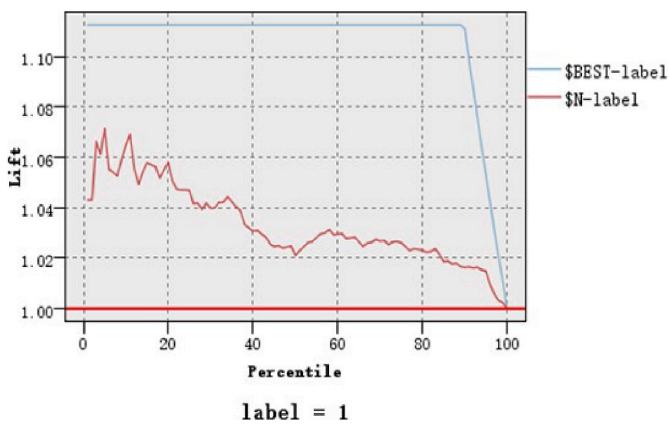

Fig. 13. The BP neural network model's lifting curve.

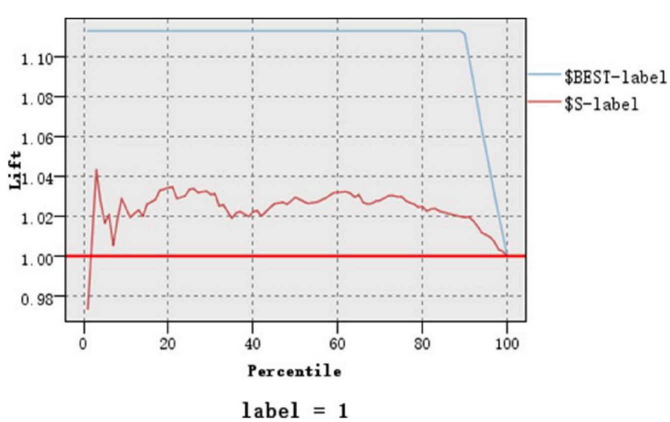

Fig. 14. The Support Vector Machine model's lifting curve.

Table 2
The value of lost consumers is classified.

Customer Category	Quantity	Rate%
E_1N_1	836	58.7
E_1N_2	370	25.5
E_2N_1	150	9.8
E_2N_2	120	7.9

considerably lower the rate of customer churn when lost customer retention tactics are utilised. However, due of their great worth and minimal likelihood of being lost, it is preferable to try to get them back. According to previous study, older customers are more likely to be financially valued and more regular than new ones. Calculating the value of lost customers demonstrates that as the volume and frequency of transactions increase, the risk of client turnover drastically declines. Owners of e-commerce businesses should concentrate on the two factors of purchase volume and frequency, create a strategy for customer retention, effectively lower customer churn, and achieve sustainable company development.

3. Conclusion

Multiple marketing studies have shown that retaining existing consumers is the greatest option for sustaining a B2C E-Commerce business, as penetration of the competition's client base is nearly five times more expensive than retention. Furthermore, a new customer must be considered valuable after a longer period and a greater number of transactions. According to a study conducted by Bain & Co, a. 5% increment in client maintenance can bring about a 25% expansion in benefit. Another figure shows that dynamic clients accomplish more Business than new shoppers in B2C E-Commerce. The likelihood of Business from dynamic clients is 60%. These variables underline the significance of holding clients and reducing churn. This study provides a framework that uses machine learning methods, specifically support vector machines, to detect probable client attrition. The huge measure of information accessible to E-Commerce organizations can be mined to find covered standards of conduct, and any deviation from expected patterns or examples can be seen as a potential client churn. Personalized retention techniques are also developed using hybrid recommendation strategies. The following steps are planned, evaluation of recommendation efficiency and automatic feedback to improve the SVM model. In addition, ensemble methods are compared to the SVM model to see if they may be used to increase the churn detection process.

Funding statement

The authors receive no external funding for this research

CRediT authorship contribution statement

Shobana J: Supervision, Writing – review & editing. **Ch. Gangadhar:** Methodology. **Rakesh Kumar Arora:** Data curation. **P.N. Renjith:** Supervision, Writing – review & editing. **J. Bamini:** Conceptualization. **Yugendra devidas Chincholkar:** Validation, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] K. Coussement, K.W. De Bock, Customer churn prediction in the online gambling industry: the beneficial effect of ensemble learning, J. Bus. Res. 66 (9) (2013) 1629–1636.

[2] P. Berger, M. Kompan, User modeling for churn prediction in E-commerce, IEEE Intell. Syst. 34 (2) (2019) 44–52.

[3] Ponnan Suresh, J Robert Theivadas, V.S. HemaKumar, Daniel Einarson, Driver monitoring and passenger interaction system using wearable device in intelligent vehicle, Comput. Electr. Eng. 103 (2022), 108323, https://doi.org/10.1016/j.compeleceng.2022.108323. ISSN 0045-7906.

[4] X. Wu, S. Meng, E-commerce customer churn prediction based on improved SMOTE and AdaBoost, in: 2016 13th International Conference on Service Systems and Service Management (ICSSSM), IEEE, 2016, June, pp. 1–5.

[5] S. Markkandan, R. Logeshwaran, N. Venkateswaran, Analysis of precoder decomposition algorithms for MIMO system design, IETE J. Res. (2021), https://doi.org/10.1080/03772063.2021.1920848.

[6] G. Klimantaviciute, Customer churn prediction in E-commerce industry, J. Mach. Learn. Res. 1 (2021) 1–14.

[7] (a) Q. Yanfang, L. Chen, Research on E-commerce user churn prediction based on logistic regression, in: 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), IEEE, 2017, December, pp. 87–91;
(b) Y. Zhuang, Research on E- commerce customer churn prediction based on

improved value model and XG-boost algorithm, Manag. Sci. Eng. 12 (3) (2018) 51–56.

[8] S. Renjith, An integrated framework to recommend personalized retention actions to control B2C E-commerce customer churn, 2015. Volume-27 Number-3, pp 152-159 arXiv preprint arXiv:1511.06975.

[9] E.E. Grandón, P.E. Ramírez-Correa, J.S.L. Orrego, Modelo de Aplicaciones e-Business en Grandes Empresas: una Validación Empírica, Interciencia 44 (4) (2019) 210–217.

[10] X.Q. Wu, L. Zhang, S.L. Tian, L. Wu, Scenario based e-commerce recommendation algorithm based on customer interest in Internet of things environment, Electron. Commer. Res. (2019) 1–17.

[11] Ponnan Suresh, J Robert Theivadas, V.S. HemaKumar, Daniel Einarson, Driver monitoring and passenger interaction system using wearable device in intelligent vehicle, Comput. Electr. Eng. 103 (2022), 108323, https://doi.org/10.1016/j.compeleceng.2022.108323. ISSN 0045-7906.

[12] Z. Pei, R. Yan, Cooperative behavior and information sharing in the e-commerce age, Ind. Market. Manag. 76 (2019) 12–22.

[13] V. Urbancokova, M. Kompan, Z. Trebulova, M. Bielikova, BEHAVIOR-BASED customer demography prediction in e-commerce, J. Electron. Commer. Res. 21 (2) (2020) 96–112.

[14] T.T. Leonid, R. Jayaparvathy, Classification of elephant sounds using parallel convolutional neural network, INTELLIGENT AUTOMATION AND SOFT COMPUTING 32 (3) (2022) 1415–1426.

[15] X. Yu, S. Guo, J. Guo, X. Huang, An extended support vector machine forecasting framework for customer churn in e-commerce, Expert Syst. Appl. 38 (3) (2011) 1425–1430.

[16] N. Glady, B. Baesens, C. Croux, Modeling churn using customer lifetime value, Eur. J. Oper. Res. 197 (1) (2009) 402–411.

[17] K.A. Amuda, A.B. Adeyemo, Customers Churn Prediction in Financial Institution Using Artificial Neural Network, 2019 arXiv preprint arXiv:1912.11346.

[18] T. Ferreira, I. Pedrosa, J. Bernardino, Business intelligence for e-commerce: survey and research directions, in: World Conference on Information Systems and Technologies, Springer, Cham, 2017, April, pp. 215–225.

[19] J. Yadav, B. Mallick, Web mining: characteristics and application in ecommerce, Intl. J. of IJECSE 1 (4) (2011).

[20] A. Bologa, R. Bologa, Business intelligence using software agents, Database Syst. J. 2 (4) (2011) 31–42.

[21] Dilip Singh Sisodia, Somdutta Vishwakarma, Abinash Pujahari, Evaluation of machine learning models for employee churn prediction, in: International Conference on Inventive Computing and Informatics, 2017.