ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 11 (2024) No. 6, pp. 460-471
DOI:10.33168/JLISS.2024.0627

Evaluating Machine Learning Techniques for Predicting Customer Churn in E-Commerce: A Comparative Analysis

Layth Almahadeen

Department of Financial and Administrative Sciences, Al-Balqa Applied University, Salt, Jordan
layth.mahadeen@bau.edu.jo

Abstract. This study intends to demonstrate different machine-learning techniques for business intelligence applications. These techniques, identified as clustering, decision trees, naive Bayes, support vector machines (SVM), and logistic regression, provide their utility for e-commerce customer churn forecasts, as one of the highlighted methodological procedures, thematic analysis was used against prevailing literature to recognise strengths, weaknesses and probable implementations of every technique, specifically in business intelligence. Integrating machine learning techniques for predicting customer churn rate can be considered a contributory aspect of this study. Identification of differences can be useful for e-commerce owners to avoid undesired issues in business operations. These themes are meant to fulfil the aim of this research, to contrast different machine learning methods. Further, the study identified the significance of technique choice as dependent on e-commerce dataset aspects, exchanges within interpretable capability, precision, and computational efficacy. The overall findings, divided into specific themes, contribute to the identification of the benefits of each technique. The findings, through comparison, added specific inference to the flexibility, simplification and suitability of decision trees for categorical data management. The research study guides e-commerce in implementing machine learning as a practical implementation for diminishing business customer churn. Additionally, the research provides a prospective scope for future studies for observational appraisal regarding the discussed techniques applying realistic e-commerce data to validate their efficiency.

Keywords: Machine learning, business intelligence, customer churn, e-commerce

1. Introduction

The shopping habits of society are changing, and it is an opportunity for modern firms. It is common among e-commerce websites to generate adequate revenue with the help of important steps and methods (Guthrie et al. 2021). A high satisfaction level among customers can lead towards a lower customer churn rate. According to Germain (2023), an estimated 42% of B2C firms report having over 3% monthly churn rates, with 16% of businesses having over 4% churn rates and an overall average being near 6% for several companies. According to Angelopoulos et al. (2019), machine learning can be addressed as a useful intervention method that can be useful during decision-making. In other words, implementing machine learning techniques can be helpful for business owners while undertaking crucial decisions.

The emergence of machine learning as a disruptive innovation from Industry 4.0 might assist modern businesses in identifying customer demands and minimising the negative impact of customer churn rate. As opined by Kumar & Zymbler (2019), machine learning approaches can be useful while analysing large sets of databases. Hence, it can be useful for e-commerce websites to integrate appropriate machine learning methods, allowing for higher success. Companies need to prioritise customer retention, thereby tackling the risks of churn rate, making it urgent for e-commerce to identify the implementation of suitable machine-learning techniques for churn predictions.

Training machine learning for generating predictive models for churn rate forecasts through vast datasets also requires understanding flexible and suitable techniques for application. As per Lalwani et al. (2022), machine-learning techniques, including decision trees, logistics regression, support vector machine, and naive Bayes, are applicable to predict customer churn rates. It can further allow e-commerce owners to adopt useful tactics that can keep customers engaged and loyal towards the company. The study focuses on the objective of comparative analysis of different machine learning techniques. Comparative analysis of customer churn-oriented machine learning techniques is an essential subject for e-commerce analytics and business intelligence literature. This study investigates the merits and possible deficiencies of different churn prediction machine learning techniques in the context of business intelligence and e-commerce analytics.

2. Methodology

The completion of this study can be linked to implementing a justified methodological approach to secondary qualitative research. A secondary qualitative method will be used to compare different techniques for churn prediction based on existing literature. Several studies have identified the thematic analysis process as a part of the qualitative data analysis approach (Campbell et al. 2021). Information from authentic sources such as journals and articles was evaluated during the initial phase. Generating themes for this study were derived from the assessed literature to attain the overall aim of this study. The evaluation literature, with the aid of recognised keywords related to e-commerce, customer churn and machine learning techniques, is beneficial in developing critical observations. Critical scrutiny of existing literature provided further details regarding different machine learning churning prediction techniques. Potential gaps from thematic analysis intriguing possible evaluation bias were mitigated by focusing on both strengths and drawbacks of the five explored techniques for churn predictive model and business intelligence implementation.

3. Finding

Table 1: Machine Learning Techniques for comparative analysis

	Clustering method	**Support Vector Method (SVM)**	**Logistics Regression**	**Naïve Bayes**	**Decision Tree**
Predictive accuracy	Predictability through identification of similar data and accuracy of algorithmic grouping	Increasing accuracy through Maximising margin separation between classes and detection of subspace lies	High predictive power through predictor and dependent variable correlation	Probability classifier through historical data supported accuracy prediction	High accuracy rate with dependency on data
Interpretability	Optimisation through similarity in data points and generating cluster hierarchy	Model error generalisation, classification, regression, transactions	Dependent and independent variable relationship interpretation	Independence of features	Generation of simple decision rules
Scalability	Large dataset scalability	High dimensional space feature data mapping	Linear data presentation through	Historical data-dependent scalability through recognised features	Scalability through variable relations from historical data sea
Ease of use	Limited to selected clusters	Reducing computational complexity	Easier training facilities	Justification of prediction made within business operations	Easy readability, flexibility for ease of use

3.1. Clustering method

The clustering method can be considered a useful approach that adequately illustrates the beneficial aspects of machine learning approaches. According to Ezugwu et al. (2022), clustering methods include grouping individual data sets per their features. In other words, this method often identifies specific groups of datasets based on similar qualities and attributes. It can also be addressed as a data mining tool that can be used for

data analysis purposes. The name "Automated clustering algorithm" can be mentioned in this context, and it can formulate clustering without pre-existing knowledge regarding the data set. Different types of data clustering techniques can be observed among researchers meant to solve clustering issues. In other words, it can be considered a critical form of machine learning aimed towards providing diverse solutions through advanced technology.

The induction idea of clustering has been based on Kohonen's self-organising algorithms, and the function and purpose of optimising are based on the similarity and distance between the clusters (Luaces et al. 1999). The main principle of clustering is to discover the similarity between data points and establish a cluster hierarchy. The principle of this method is to recover the model and use it to determine the data points that help build clusters of similar data points (Ezugwu et al., 2022). Study shows that several assumptions need to be made for k-means clustering, including decision-making, data pre-processing, scaling and the relevance of test results for clustering.

The assumptions made for data analytics are often contained in data cleaning and pre-processing. Assumptions also need to be made to maintain the standard guidelines and improve the analytical reuse of the data (Wenzlick et al. 2021). There are several advantages of clustering, as the clusters are not assumed to be globular. Using the clustering algorithm also makes them more efficient in scaling large datasets. This machine learning algorithm could help design patterns to deliver better services (Ezugwu et al. 2022). Limitations include a lack of capturing discriminative information in low dimensional subspace and vulnerability to noise and optimisation. The user also needs to choose the number of clusters acting as a limitation to effective implementation. The study by (Addagarla & Amalanathan 2020) shows that the medical profession often uses clustering to cluster medical images for various analyses through artificial intelligence.

The market demand in the modern era is changing regularly, and it can influence the performance level of a firm. Modern technologies are useful for firms as they can improve their operational efficiency. Performance improvement is mandatory to increase satisfaction levels among the target consumer base. In this specific matter, it can be addressed that business owners can apply clustering machine learning methods to attain a competitive edge in the market. According to the findings, the clustering machine learning technique is useful for locating similar customers (Monil et al. 2020). Modern businesses can commonly create different groups per their customer's needs. Modern e-commerce websites can use different tactics for each group and achieve high financial success. It can also be stated that modern e-commerce websites can develop ideas to fulfil changing customer demands. The clustering method is useful for showing the different needs of the customers, and it is helpful for firms.

Modern firms can increase their knowledge base through the usage of clustering methods. The decision-making process of the e-commerce owners often requires support from detailed data. It also means that e-commerce owners commonly undertake proper decisions based on identified data regarding customer preferences. Clustering methods can be helpful during the identification of similarities or patterns in large data sets, and they can be applied by business owners in the modern era to monitor similarities (AL-Najjar et al. 2022). As mentioned earlier, this method can be useful for identifying different needs of the large customer base, which can help increase customer satisfaction. Modern e-commerce owners can utilise this method to detect the patterns and demands of the different customer groups, which is essential for conducting approaches to customer satisfaction.

3.2. "Support Vector Machine" aka SVM method

SVM is another crucial app that can be presented as an advanced machine-learning technique. SVM is a new technology that functions as a supervised model for machine learning (Boateng et al. (2020). It is a

learning algorithm popular in the current era, and it can achieve good results after the analysis process. This machine learning process is useful for image detection, speech detection, and business owners. In other words, business owners can use this model to identify similar speech and images of the customers and create customer profiles. It is helpful for e-commerce websites to create customer profiles and provide personalised advertisements to attract them. The overall process of SVM takes adequate time and time is essential for modern business owners. Reasons such as these can be evaluated by the modern business owners of e-commerce websites, which can significantly impact the performance rate. Hence, modern e-commerce websites must inspect these factors and take proper measures. Several studies have identified it as a time-consuming factor which can be problematic for modern business owners. It is common for modern business owners to face competitive challenges in the market that require the adoption of necessary measurements in time.

The idea of a Support Vector Machine is built on specific principles; it understands that modern problems are highly dimensional, and the real-life data generation laws are much different from normal distribution and model building. The SVM inductive principles are mostly developed for standard contemporary data sets. The main principle is maximising the margin between the support vectors using kernel functions. This machine-learning method works by mapping data to high-dimensional feature spaces. Researchers consider SVM the inductive principle directed at learning from finite training data sets (Kecman, 2005). Like the other machine learning techniques, SVM is also built on some assumptions; the best line maximises the margin between both classes.

One of the assumptions is that the classes to be discriminated against lie in a subspace (Osuna, Freund, & Girosi, 1997). One of the main advantages or strengths of SVM is that this method works well when there is a clear margin of separation between the different classes. This method works well in high-dimensional spaces. Another significant advantage is its ability to use a whole set of data for training and generalising the error of the models. On the other hand, the study's limitations and disadvantages lie in using the pessimistic behaviour of the Maximal Discrepancy method and reducing the computational complexity (Anguita et al. 2010). This machine learning method solves classification, regression tasks, and e-commerce transactions. Additionally, SVM algorithms can find document similarity and help the organisation with the product review category (Hadju & Jayadi, 2021).

SVM is also helpful in detecting customer preferences, which can benefit e-commerce owners. According to Hong et al. (2020), SVM is useful in detecting customer satisfaction, which indicates a firm's performance level. In other terms, firms can invest in this type of machine learning approach that can allow them to improve their operational activities. It can also enable decision-makers of modern firms to identify valuable options that can be applied to ensure a high-performance rate. It can also be linked with a high satisfaction rate among consumers, which can be considered an outcome of the high performance of e-commerce websites. It can also be stated that e-commerce owners can make proper decisions linked with a high-performance rate. It can further allow them to ensure satisfaction among the target consumer group linked with the financial performance rate.

The role of SVM in detecting patterns and crucial numbers cannot be ignored, as they are linked with customer satisfaction rates. Ly et al. (2022) have also discussed the usage of SVM in classifying binary numbers that can be addressed as advanced machine learning practices. In other words, using SVM techniques can help decrease the number of misclassifications. It can further benefit churn prediction, allowing e-commerce websites to improve their financial performance rate. The high accuracy rate can be mentioned here and can be expected from using this technique during churn prediction. The high dependency on the hyper-parameters can also be linked with its misclassification, creating confusion during decision-making. In other words, misclassification can lead to wrongful decisions and reduce the scope for

high customer satisfaction.

3.3. Logistic Regression for Understanding and Prediction of Customer Churn

Among the different machine learning techniques, logistic regression is one of the most effective measures. Logistic regression can be understood as a statistical method that can be utilised for building machine-learning models. The primary purpose of this model is to elaborate on the relations between the dependent and independent variables of the research. Studies have shown that in the past few years, different churn models have been introduced with the help of different techniques. Many organisations today need to adopt machine learning techniques that can predict customer churn. Logistic Regression and Logic Boost are other prediction models used for filtering and data cleaning. Logistic regression as a machine learning algorithm should not be mistaken for the black box model, as the findings can be binary, multinomial or ordinal (Jain, Khunteta & Srivastava, 2020). Prediction-making in the context of real-valued inputs in the machine learning process can be considered one of the most crucial aspects of the present scenario. A value greater than 0.5 is accumulated as class 0, while other values fall into class 1. In this regard, class y has been identified to be effective in prediction according to the range mentioned.

The idea of logistic regression has been based on underlying principles and probabilities and the nature of the log curve. Researchers suggest that the only assumption that can be made for logistic regression is that the logit transformation is generally depicted to be linear. The dependent variables are dichotomous because the resultant logarithmic curve has no outliners (Healy, 2006). The advantages of logistic regression are that it evenly portrays the increased or decreased likelihood of event outcome occurrences. When the odd ratio is reported to be less than the decreased likelihood, the odds ratio between the different variables is raised. One of the primary strengths of this method is its prediction of customer churn, its interpretability, and its predictive power. It is also much easier to train than using other machine learning applications. It has become incredibly important for organisations to understand customer churn; logistic regression provides this opportunity, allowing them to focus on more important parts of the business, such as profit maximisation.

Through logistic regression, profit maximisation can be achieved, and those that are most profitable towards the business can be identified through this method (Stripling et al. 2018). There are certain limitations, however, of choosing this machine learning method, the right predictor variables need to be chosen, variables that are highly correlated need to be avoided, and the regression model also assumes the relationship between predictor and dependent variables, which, if not uniform can create issues (Ranganathan, Pramesh & Aggarwal, 2017). The main application of this method is to predict customer behaviour. Study shows that with the advent of fast e-commerce development, many organisations had to be more involved in microfinance (Wang & Han, 2021). Alibaba initially led this movement, and different SMEs use logistic regression to discover their financial difficulties.

Customer Churn Prediction (CCP) has been identified as one of the most critical issues in the telecommunication industry. Hence, logistic regression can be considered one of the most effective processes in influencing algorithms and re-sampling methods. In this regard, the correlation between the variables can be accumulated using regression, a statistical analysis. Several studies have suggested that this model is useful for detecting customers' problems and creating solutions (Lalwani et al. 2022). In other words, this model can help firms make good, helpful decisions during the decision-making phase. In this context, it can be stated that the organisation gained the capability of accumulating consumer data, which is an essential aspect of managing retention programs. Logistics rows are considered the consumers, and the columns display preferences and attributes. For predicting the churning mechanism, the organisation will have to visualise the historical data on the consumers through these learning algorithms and then determine the likelihood of the consumer churning.

Consumers generally have many choices; therefore, improving business intelligence and predicting consumer preferences have become important. Since logistic regression is a binary classification, the churn data can predict if a certain consumer is willing to discontinue service (Devriendt, Berrevoets & Verbeke, 2021). Based on the results, the organisation can provide consumers with the highest likelihood of churning with more incentives. Incentives can be provided through discounts or promotional offers, which can help push consumers to extend their subscriptions. Overall, the algorithm chooses consumers who are most likely to churn and can be targeted through the organisation's retention campaign. Making accurate predictions is the goal of all machine-learning technologies. Logistic regression is the standard predictive model used within the industrial setting. Logistic regression has been seen to facilitate the interpretation of resulting models, and performance has been relatively good over the years.

3.4. Naïve Bayes for Customer Churn Prediction in E-commerce

The Naïve Bayes Classifier (NBC) is another machine learning algorithm organisations can use to classify different types of tasks. It is a simple probabilistic classifier that utilises Bayes theorem for making assumptions. The NBC is also known as the independent feature model. In simplistic terms, the classifier assumes that the presence or absence of a feature with a business does not affect other features and is completely unrelated. The algorithms used can also increase the prediction rates of the classifier. In their studies, (Valdiviezo-Diaz et al. 2019) have shown that NBC allows understanding and justification of the different predictions made within the organisation's business process. Furthermore, it is a classification algorithm of the Bayes theorem applied to achieve different “naive” assumptions.

The focus of NBC is to classify a certain score, P(C|X), that is proportional to the posterior probability and helps maximise the classification score of the probability. While several different forms of NBC have been proposed over the years, they differ based on the assumptions about their distribution. The method can classify or predict new consumer ratings based on their historical data with the organisation. Yulianti & Saifudin (2020) in their research make it clear that NBC is a machine-learning method that can only provide accurate results when enough data sets are provided to the system. Since organisations need to improve customer services to retain consumers, knowing the consumers who are likely to leave is beneficial for the organisation. Using NBC can, therefore, help retain consumers and generate profits. The datasets in this context are highly important because flawed datasets can lead to redundancy or irrelevancy of the research findings.

Customer churn generally occurs due to a lack of communication, dissatisfaction, and a general preference for other products. Building long-term relationships with consumers can help prevent them from going into other organisations. Data mining is important for these organisations to find patterns in consumer behaviour. As a result, NBC is an effective classification which has been built to provide independent results and structure. Research has found that the predictions made by NBC can further be improved with the help of SMOTE. Naive Bayes has a success rate of 47.10%, and the accuracy level rises to 78.15% when combined with SMOTE (Safitri & Muslim, 2020). The potential of Naive Bayes classification has also increased with the addition of Genetic Algorithms, which can help organisations with their business intelligence applications.

3.5. Training Decision Tree Model for Customer Churn Prediction

Decision trees are recognised as a machine learning technique; these algorithms are tree-shaped and represent sets of decisions that can be used for generating classification rules for a particular data set. With the help of this model, organisations can divide large data sets into smaller records by applying simple decision rules. The decision tree is sometimes also called a classification tree or regression tree. Class labels represent the leaves of the tree, while the branches become features that lead to the labels. While decision

trees generally cannot illustrate complicated and non-linear variable relations, the accuracy of this algorithm is relatively high depending on the data used for prediction purposes. In their research on Decision Tree, Kim & Lee (2022) found that the algorithm has a maximum prediction accuracy of 90%. The decision tree, in particular, is preferred due to its easy-to-use nature and high accuracy of the results it provides. Decision trees can be used to predict customers' churning through analysis of their e-commerce data.

Decision tree values are binary, making them optimal for machine learning algorithms, if a consumer has made a purchase from an organisation or influencer, it can be considered as a churn, the opposite of which will make them a loyal consumer towards the brand. Decision Tree algorithm has the highest possibility of making predictions, it is however combined with other models instead of using as a singular model to increase its low performance and make the algorithm more robust (Kim & Lee, 2022). In the modern world, customer churn has become a considerable concern due to the highly competitive nature of the services that are provided, the fluctuating nature of the consumer in moving from one provider to another can be explained with the help of this model (Ahmad, Jafar & Aljoumaa, 2019). Implementing a decision tree at an early stage within the organisation can not only make the organisation more competitive but also open additional sources of revenue, machine learning techniques such as decision trees are highly effective for these scenarios.

With the help of a decision tree, the collected data of an organisation can be sliced into various branch-like segments. The findings of the tree are easier to read, these advantages make the explanation for the models simple. Even though the other machine learning algorithms can also produce accurate results, the decision tree can be trained to predict neural network predictions (Rahman & Kumar, 2022). Another significant factor is the correlation between the target variables and predictor variables is the high degree of non-linearity between the factors.

4. Discussion

The present research has discussed four different machine-learning techniques, namely Clustering, Support Vector Machine (SVM), Logistic Regression, Naïve Bayes and Decision Tree. The first focus of the research has been provided on clustering, which can be understood as an algorithm that analyses different groups of objects by bringing them together. Unlike the other methods analysed in this section, clustering cannot be completely automated, the present research has argued that with the help of this method, consumers can be segmented based on their interests and preferences, making this process important as a customer churn predictor. The second technique is SVM, the data in this method are plotted in a n-dimensional space, and the classifications of consumers in this context are done on a suitable hyper-plane (Vats & Sharma, 2020). In SVM there is a maximum margin classifier and a soft margin. While clustering helps in assorting consumers in certain categories, SVM helps by predicting the customer satisfaction aspect. The hyper lane identification capability of SVM makes it useful for classifying binary numbers.

Logistic regression is one of the more popular machine-learning algorithms, and even though the name regression comes up, it is in general a classification model the purpose of which is to frame the binary output model (Dhamodharavadhani &Rathipriya, 2020). There are several benefits of this model, using this model is easy and fast. The parameters used in this technique explain the direction and intensity of the significance, the regression model can also be applied in multiclass classifications. There are visible differences between logistic regression and SVM, while logistic regression is only able to handle linear solutions, SVM deals with nonlinear solutions. SVM generally performs better due to its maximum-margin solutions. The use of logistic regression can help address the complex problems which consumers face, the rows and columns within the logistics show the preferences of the consumers. Logistic regression can improve business intelligence and predict consumers that are most likely to churn.

The research has also emphasised the Naive Bayes Classifier as a supervised machine learning algorithm. The algorithm is built on the Bayes Theorem and is used to make independent assumptions about consumers and their preferences, the main assumption being that all of the features that are provided by an organisation or a service provider are unrelated (Li & Li, 2020). The findings of the study with the availability of enough data, the classifier can predict accurate results. With the help of NBC, organisations will be able to improve their frameworks and raise the accuracy levels. The study also shares how the potential of the NBC can be increased with the addition of a Genetic Algorithm. The final technique discussed in the research and perhaps the most successful is a decision tree. The concept of decision tree is based on an algorithm which is used to solve regression and classification problems (Charbuty & Abdulazeez, 2021). The idea of the decision tree is derived from independent variables, as a result, each of the nodes helps with the prediction of a certain output. The advantages of using it for predicting customer churn are that assumptions are not made on the distribution of data, and proper understanding is provided of the predictions that are made. Further significance of the findings imply the underlying scope of empirical verification of the range of comparative merit of each technique's accuracy, scalability, interpretability and ease of use through practical e-commerce business intelligence and analytics.

5. Conclusion

As concluding remarks for the research, though, all five techniques compared through evaluation have their advantages, the decision tree has been identified as one of the best techniques that can help with e-commerce customer churn. Decision trees are generally better in categorical and collinear data. Decision trees also have better flexibility and ease of readability than others. E-commerce analytics through machine learning is based on exchanges of interpretability, accuracy, scalability and user-friendly nature of the tools. The inferences make this study a contributor to the subjects of machine learning and business intelligence for customer churn prediction for e-commerce. This study being limited to only the use of secondary data, facilitates prospects for future researches implement empirical validation through practical e-commerce datasets.

References

Addagarla, S. K., & Amalanathan, A. (2020). Probabilistic unsupervised machine learning approach for a similar image recommender system for E-commerce. Symmetry, 12(11), 1783.

Ahmad, A. K., Jafar, A., &Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data, 6(1), 1-24.

AL-Najjar, D., Al-Rousan, N., & AL-Najjar, H. (2022). Machine learning to develop credit card customer churn prediction. Journal of Theoretical and Applied Electronic Commerce Research, 17(4), 1529-1542.

Angelopoulos, A., Michailidis, E. T., Nomikos, N., Trakadas, P., Hatziefremidis, A., Voliotis, S., &Zahariadis, T. (2019). Tackling faults in the industry 4.0 era—a survey of machine-learning solutions and key aspects. Sensors, 20(1), 109.

Anguita, D., Ghio, A., Greco, N., Oneto, L., & Ridella, S. (2010, July). Model selection for support vector machines: Advantages and disadvantages of the machine learning theory. In The 2010 international joint conference on neural networks (IJCNN) (pp. 1-8). IEEE.

Boateng, E. Y., Otoo, J., &Abaye, D. A. (2020). Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: a review. Journal of Data Analysis and Information Processing, 8(4), 341-357.

Campbell, K. A., Orr, E., Durepos, P., Nguyen, L., Li, L., Whitmore, C., ... & Jack, S. M. (2021). Reflexive thematic analysis for applied qualitative health research. The Qualitative Report, 26(6), 2011-2028.

Charbuty, B., &Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2(01), 20-28.

Devriendt, F., Berrevoets, J., &Verbeke, W. (2021). Why you should stop predicting customer churn and start using uplift models. Information Sciences, 548, 497-515.

Dhamodharavadhani, S., &Rathipriya, R. (2020). Enhanced logistic regression (ELR) model for big data. In Handbook of research on big data clustering and machine learning (pp. 152-176). IGI Global.

Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., & Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. Engineering Applications of Artificial Intelligence, 110, 104743.

Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., & Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. Engineering Applications of Artificial Intelligence, 110, 104743.

Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., &Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. Engineering Applications of Artificial Intelligence, 110, 104743.

Germain, J. M. (2023). Subscription commerce churn rate worldwide 2022, by product category. Retrieved on 17 March 2024, from: https://www.ecommercetimes.com/story/subscription-commerce-merchants-innovate-amid-rising-churn-rates-177730.html

Guthrie, C., Fosso-Wamba, S., & Arnaud, J. B. (2021). Online consumer resilience during a pandemic: An exploratory study of e-commerce behavior before, during and after a COVID-19 lockdown. Journal of retailing and consumer services, 61, 102570.

Hadju, S. F. N., & Jayadi, R. I. Y. A. N. T. O. (2021). Sentiment analysis of indonesian e-commerce product reviews using support vector machine based term frequency inverse document frequency. Journal of Theoretical and Applied Information Technology, 99(17), 4316-4325.

Healy, L. M. (2006). Logistic regression: An overview. Eastern Michighan College of Technology.

Hong, S. H., Kim, B., & Jung, Y. G. (2020). Correlation analysis of airline customer satisfaction using random forest with deep neural network and support vector machine model. International Journal of Internet, Broadcasting and Communication, 12(4), 26-32.

Jain, H., Khunteta, A., &Srivastava, S. (2020). Churn prediction in telecommunication using logistic regression and logit boost. Procedia Computer Science, 167, 101-112.

Kecman, V. (2005). Support vector machines–an introduction. In Support vector machines: theory and applications (pp. 1-47). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kim, S., & Lee, H. (2022). Customer churn prediction in influencer commerce: An application of decision trees. Procedia Computer Science, 199, 1332-1339.

Kumar, S., &Zymbler, M. (2019). A machine learning approach to analyze customer satisfaction from airline tweets. Journal of Big Data, 6(1), 1-16.

Lalwani, P., Mishra, M. K., Chadha, J. S., &Sethi, P. (2022). Customer churn prediction system: a machine learning approach. Computing, 1-24.

Li, Q. N., & Li, T. H. (2020). Research on the application of Naive Bayes and Support Vector Machine algorithm on exercises Classification. In Journal of Physics: Conference Series (Vol. 1437, No. 1, p. 012071). IOP Publishing.

Luaces, O., del Coz, J. J., Quevedo, J. R., Alonso, J., Ranilla, J., & Bahamonde, A. (1999, June). Autonomous clustering for machine learning. In International Work-Conference on Artificial Neural Networks (pp. 497-506). Berlin, Heidelberg: Springer Berlin Heidelberg.

Ly, T. V., & Son, D. V. T. (2022). Churn prediction in telecommunication industry using kernel Support Vector Machines. Plos one, 17(5), e0267935.

Monil, P., Darshan, P., Jecky, R., Vimarsh, C., & Bhatt, B. R. (2020). Customer segmentation using machine learning. International Journal for Research in Applied Science and Engineering Technology (IJRASET), 8(6), 2104-2108.

Osuna, E., Freund, R., & Girosi, F. (1997, September). An improved training algorithm for support vector machines. In Neural networks for signal processing VII. Proceedings of the 1997 IEEE signal processing society workshop (pp. 276-285). IEEE.

Rahman, M., & Kumar, V. (2020, November). Machine learning based customer churn prediction in banking. In 2020 4th international conference on electronics, communication and aerospace technology (ICECA) (pp. 1196-1201). IEEE.

Ranganathan, P., Pramesh, C. S., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: logistic regression. Perspectives in clinical research, 8(3), 148-151.

Safitri, A. R., & Muslim, M. A. (2020). Improved accuracy of naive bayes classifier for determination of customer churn uses smote and genetic algorithms. Journal of Soft Computing Exploration, 1(1), 70-75.

Stripling, E., vanden Broucke, S., Antonio, K., Baesens, B., & Snoeck, M. (2018). Profit maximizing logistic model for customer churn prediction using genetic algorithms. Swarm and Evolutionary Computation, 40, 116-130.

Valdiviezo-Diaz, P., Ortega, F., Cobos, E., & Lara-Cabrera, R. (2019). A collaborative filtering approach based on Naïve Bayes classifier. IEEE Access, 7, 108581-108592.

Vats, D., & Sharma, A. (2020). Dimensionality Reduction Techniques: Comparative Analysis. Journal of Computational and Theoretical Nanoscience, 17(6), 2684-2688.

Wang, P., & Han, W. (2021). Construction of a new financial E-commerce model for small and medium-sized enterprise financing based on multiple linear logistic regression. Journal of Organizational and End User Computing (JOEUC), 33(6), 1-18.

Wenzlick, M., Mamun, O., Devanathan, R., Rose, K., & Hawk, J. (2021). Data science techniques, assumptions, and challenges in alloy clustering and property prediction. Journal of Materials Engineering and Performance, 30, 823-838.

Yulianti, Y., &Saifudin, A. (2020, July). Sequential feature selection in customer churn prediction based on Naive Bayes. In IOP Conference Series: Materials Science and Engineering (Vol. 879, No. 1, p. 012090). IOP Publishing.