Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU34000

Application of Artificial Intelligence in Customer Churn Prediction: An Investigation

Xingyu Chen

Qingdao Menaul School, Qingdao, China
xingc2135@gmail.com

Abstract. Customer churn prediction has become an important task for enterprises because retaining existing customers is usually much less expensive than acquiring new ones. With the rapid growth of customer behavioral data, artificial intelligence and machine learning technologies have been widely applied to identify customers who are likely to leave a service or platform. This paper reviews recent applications of artificial intelligence in customer churn prediction, including traditional machine learning methods such as logistic regression, decision trees, and random forests, as well as deep learning models including artificial neural networks (ANNs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), and attention-based methods. The advantages and limitations of these approaches are discussed from the perspectives of prediction accuracy, interpretability, privacy protection, and cross-domain adaptability. In addition, this paper summarizes several important future research directions, including expert systems, federated learning, and domain adaptation techniques. These methods are expected to improve the reliability, privacy, and practical application ability of future customer churn prediction systems in different industries.

Keywords: Customer churn prediction, Machine learning, Explainable AI, Domain adaptation

1. Introduction

Now, customer retention has begun to affect the profits of all types of enterprises in the digital economy. Research shows that it is 5 to 25 times more expensive to acquire new customers than to retain existing ones; therefore, early identification of at-risk customers, often referred to as "customer churns" or "churners", is required by all sectors of the economy to protect business interests [1]. Customer churn refers to a particular customer reducing their use of a certain service, such as decreasing the frequency of purchases or a lower frequency of using the platform; they may not be using the service at all. A general operating definition of customer churn in the literature is that a customer has not performed any substantial activities or fails to exhibit the expected contact patterns for a period of about 30 days.

With the spread of digital touchpoints, a large amount of customer behaviour data has been generated, and enterprises are now trying to use this data for the prediction and prevention of customer churn. The old rule-based system and manual division method cannot cope with the high

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.GU34000

dimensionality, heterogeneity and time-variance of new customer data [2]. Due to this limitation, Artificial Intelligence (AI) and machine learning (ML) techniques have been introduced to the construction of CRM systems, and now the company will use the collected data to proactively intervene rather than respond to complaints.

The development of artificial intelligence in customer analysis has had various stages. In the early days, statistics and some simple machine learning models, such as logistic regression, decision trees, and random forests, were used to build models; they were easy to understand but not very powerful in prediction [3]. With the start of deep learning (DL) [4], some excellent types of artificial neural networks (ANNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs) were used to study all sorts of data on the behaviour of consumers and other complex patterns. Recently, an ensemble and hybrid method combining different kinds of deep learning and attention mechanisms along with optimisers has also been used to solve this problem in recent years, and good prediction results have been achieved.

The current situation of research on the application of artificial intelligence in customer churn prediction with traditional machine learning and deep learning methods will be introduced in this paper. The organisation of this paper is as follows: Section 2 is about the traditional Machine Learning and Deep Learning technologies in this study; Section 3 describes problems that still need to be solved in this field, such as explainability, privacy and generaliseability; Section 4 proposes some directions for development of expert systems, federated learning, domain adaptation, etc.; Section 5 summarises the present results of the study and provides implications for future research and practice.

2. Application of artificial intelligence in customer churn prediction

2.1. Traditional machine learning algorithms

2.1.1. Logistic Regression

Logistic Regression (LR) is an old and simple method that has been used many times in predicting customer churn. LR is a general linear model that takes the logit of the odds of a binary outcome as a linear function of the input features to predict the probability of customer churn. It is very simple, and in a particular case, an easily understandable method was needed. For example, in the problem of telecommunications churn prediction, the LR model has reached a baseline accuracy of over 75% and can provide an intuitive understanding of how much each factor influences the probability of a customer leaving [5]. According to the probability distribution of the model, a risk level will be determined, and then the company will focus its intervention on the areas with a high probability. LR thinks that the features are linearly correlated with the log-odds of the outcome, so it cannot capture all types of non-linear relationships among the data of customer behaviour.

2.1.2. Decision tree

Decision trees (DTs) are easy to understand and can be divided into several homogenous regions in the feature space by recursively applying information gain or Gini impurity criteria [6]. DTs have an inner hierarchy and can solve the problems of non-linearity and feature interaction in customer segmentation and churn prediction. DTs have been used in practice to determine the main reasons for churn in the telecommunications industry, such as contract length, monthly fees and service usage. DTs have the shape of flowcharts, so they are easier for people to understand and support

DOI: 10.54254/2755-2721/2026.GU34000

business decisions. A single DT may be prone to overfitting due to the large number of customer attributes and has relatively high variance; thus, it is very sensitive to fluctuations in the data.

2.1.3. Random Forest

To solve the problem of a single decision tree, multiple decorrelated trees can be built by means of ensemble learning techniques, and in this process, bootstrap aggregation (bagging) and random feature subset selection are used to construct a Random Forest (RF). Therefore, it is possible to reduce overfitting and improve both the prediction accuracy and stability. Generally speaking, RF has shown good results compared with single DTs and logistic regression for the problem of customer churn prediction; many studies have shown that it can improve the accuracy of various industry datasets by 8%-15% , and the built-in feature importance indicators of the algorithm, such as mean decrease in impurity or permutation importance, can also be used to determine which customer attributes are more important [7]. Recently, comparative studies have shown that RF can reach the same accuracy as more complicated deep learning models, and it is also easy to interpret and relatively light in terms of computation; thus, it is often used in many enterprises.

2.2. Deep learning algorithms

2.2.1. Artificial Neural Networks

Artificial Neural Networks (ANNs) are the basic form of deep learning, and many connected neurons are used to learn hierarchical feature representations through backpropagation ; for the purpose of predicting customer churn, a multi-layer perceptron (MLP) has been employed to model the complex non-linear relationship between customer attributes and the occurrence of churn [4]. Some studies have found that a well-tuned and optimised ANN can perform about 5%-10% better than Random Forests in the handling of large-scale, high-dimensional customer data compared to traditional machine learning methods , and due to the flexibility of the ANN architecture, all kinds of features, such as demographic information, behaviour and transaction data, can be processed after proper encoding and normalisation. It is necessary to reasonably adjust the hyperparameters of the ANN, and if regularisation methods such as dropout and early stopping are not used, there will be a risk of overfitting [8, 9].

2.2.2. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) were first used in image processing, and then they have been applied to the problem of customer churn prediction by treating sequences of behaviour data as one-dimensional signals. CNN can learn the local patterns and hierarchical features of the sequence of customer behaviour, such as the periodicity of purchase behaviour and temporal clustering of service use. Together with a recurrent network, CNNs can be applied to solve the problems of space and time in customer data and improve the accuracy of prediction [10]. Liu and others have put forward a fused attention deep learning model (AttnBLSTM-CNN) that combines bidirectional LSTMs and CNNs with an attention mechanism, and it has been found that such hybrid architectures can improve the accuracy of churn detection by focusing on important customer behaviour patterns. Attention helps us see how important different times and attributes are in a particular problem through weightings.

DOI: 10.54254/2755-2721/2026.GU34000

2.2.3. Recurrent Neural Networks and attention mechanisms

LSTM and GRU are types of RNNs that have been employed to model the sequence of customer behaviour well. RNNs have a state and, therefore, can grasp the time sequence of customer engagement rather than the simple feedforward networks. LSTMs have been used to solve the vanishing gradient problem of traditional RNNs and can better handle long-term dependencies in customer lifecycle data by means of different gating mechanisms [11]. Recently, attention mechanisms have been added to RNNs, and now the model can weigh different historical interactions differently in prediction. Jajam and Zhao put forward an ensemble model based on stacked bidirectional LSTMs (SBLSTM) and RNNs with arithmetic optimisation; it achieved a high accuracy of 97.89% in the insurance domain churn prediction. At the same time, Zhao et al.[12] developed a multi-head self-attention-based nested LSTM classifier that is good at distinguishing between medium-value and high-value customers in the telecom industry [13].

2.2.4. Ensemble and hybrid deep learning methods

Recently, many deep learning methods have been combined in an ensemble or hybrid way to solve the problem of churn prediction. The above ways can make full use of the strengths of different models; that is, LSTMs are suitable for sequence modelling, CNNs can recognise patterns, and metaheuristic algorithms are employed for optimisation; thus, a combination of them has shown good prediction results [14]. In 2024, Van-Hieu and others proposed a stacked deep learning ensemble for banking churn prediction that combined three different neural networks at Level 0 and used logistic regression for the final prediction at Level 1; it achieved an accuracy of 96.60%, a precision of 90.26%, and a recall of 91.91%. Many models have been released to enhance the accuracy of a single model and reduce the problem of overfitting caused by fitting some data. The success of the hybrid approach shows that, facing the problem of customer churn, many methods must be employed [15].

3. Discussion

3.1. Challenges

3.1.1. Interpretability

The prediction performance of the deep learning model is good, but due to the "black-box" problem, it cannot be used in customer service. To gain the trust of the business side and decision-makers, it is necessary to explain why a model makes a certain prediction openly and formulate corresponding responses. Many neural networks are too opaque and thus not suitable for the use cases of regulated industries that need to be accountable and transparent. In light of the above problem, explainable artificial intelligence (XAI) methods such as SHapley Additive exPlanations (SHAP) and LIME (Local Interpretable Model-agnostic Explanations) have gradually been introduced into the churn prediction system [16, 17]. Based on cooperative game theory, SHAP is a general feature importance method that assigns a value to every feature for all possible combinations of features to explain a model's prediction. LIME is used to explain the individual predictions of a model by fitting a simple local model to it. Research abroad in the field of clinical and customer analytics has found that SHAP is relatively stable and generalizable in terms of explanations; it can identify the most prominent risks in the whole model better, and LIME can provide a brief explanation for individual

DOI: 10.54254/2755-2721/2026.GU34000

cases. Adding these XAI methods to the churn prediction process can help the company consider how to act based on both the prediction results and their reasons, meet regulations, etc.

3.1.2. Privacy

The use of artificial intelligence in customer analysis has aroused people's concern about privacy because these models need to obtain a large amount of personal and behavioural information of individuals. The traditional centralised machine learning method needs to collect all the customer data from various sources in one place, which will be difficult to meet the privacy laws of the EU and California [18]. Federated Learning (FL) is a very good way to train models in distributed data silos while not sharing the raw data [19]. In cross-silo federated learning, each silo is a company or other organisation with a large amount of customer data, and in order to formally protect the privacy of this data, differential privacy (DP) technology is used. Recently, new DP frameworks have been released to protect the data of individual users scattered in many places and personalized DP methods have been developed that meet the different privacy needs for different kinds of records. With the application of collective intelligence, the company will be better able to predict employee turnover under the rules of privacy protection and other regulations.

3.1.3. Application

There are many problems in the actual application of AI-driven churn prediction systems, such as data heterogeneity, concept drift and the cold-start problem [20]. The behaviour of customers in different industries, areas and groups of people varies significantly; therefore, a customised model is needed. Changes in the environment and the needs of the market and consumers are also reasons for concept drift; thus, the accuracy of the model will gradually decrease over time. The problem of the cold start is that new customers or new areas do not have historical data, so collaborative filtering and supervised learning cannot be used. Domain adaptation and transfer learning are used to make good use of the knowledge contained in the data-rich source domain and thus improve prediction accuracy on the target domain with little target domain data. Cross-domain recommendation methods are based on embeddings and mappings to directly learn transfer functions between domains, can solve the cold-start problem, but if there is not enough overlapping user data, they may not perform well in the general case.

3.2. Future development

3.2.1. Expert systems

In recent years, AI-driven churn prediction and expert systems have begun to be applied in practice together. An expert system is made up of two parts, machine-learning models and rules; it stores knowledge from a specific area and can offer various forms of decision assistance. Based on business rules, regulations and other conditions, as well as internal information from the organization, expert systems can provide some suggestions for the company's work. In the future, some research will be conducted on how to combine deep learning models with symbols to improve symbolic logic, and thus companies will be able to use both data analysis and expert judgment in their plans for customer retention.

DOI: 10.54254/2755-2721/2026.GU34000

3.2.2. Federated learning

Federated Learning will probably be one of the first means to protect the privacy of customer data in multi-party cooperation. In the future, the further study will change the way of client selection to make the model perform better in different environments and with various data distributions; this paper will also strengthen privacy protection by achieving a good trade-off between privacy and utility and apply personalised federated learning according to different scenarios of local data. In order to improve the guarantee of privacy and security for the collaborative churn prediction system in the future, more efforts will be made to integrate federated learning with differential privacy, secure multi-party computation and blockchain technology, etc. With the development of these technologies, enterprises will be able to build strong, private customer analytics ecosystems to make use of collective intelligence without violating people's privacy.

3.2.3. Domain adaptation

To solve the problem of data scarcity and cross-domain generalisation, more and more domain adaptation and transfer learning methods will be used in customer analytics. In the future, multimodal domain adaptation methods will be further developed, and data from various sources such as text, images and behaviour will be employed to promote cross-domain knowledge transfer. Sharpness-aware Minimisation is a good way to improve the generalisation ability of cross-domain models that have been trained on a small number of overlapping user data by optimising both the recommendation loss and the sharpness of the loss. With the development of unsupervised and semi-supervised domain adaptation methods, the need for a large amount of labelled data in the target domain will be reduced, and churn prediction for emerging markets and niche industries will be realised more easily.

4. Conclusion

This paper reviews the use of artificial intelligence technology for identifying customer churn in recent years and traces the development process of various machine learning and deep learning methods, etc. At present, the older models of Logistic Regression, Decision Trees and Random Forests are still used frequently because they are simple to operate and have relatively easy calculations; at the same time, Deep Learning models, such as ANNs, CNNs, LSTMs, attention-based models, etc., have shown good predictive performance and are now being applied to learn complex temporal and non-linear patterns in customer data.

There are problems with the interpretability, privacy protection and cross-domain application of the field. SHAP and LIME are some types of explainable AI technology that can help us improve the transparency of models and understand how these models make predictions. Differential privacy combined with federated learning is a general method to solve the issue of customer information privacy. Domain adaptation and transfer learning are to address the problems of data scarcity and cold start, so a churn prediction system can be deployed in many different business environments.

In the future, it is expected that many expert systems and new federated learning protocols and methods for multi-modal domain adaptation will be used to build the next-generation AI-driven customer retention platform. These changes will help the company know about a customer's departure earlier in their life, so it can make a suitable response quickly and increase the customer's total contribution to the company over time. Given that this study covers various disciplines such as computer science, management science and ethics, it is necessary to strengthen cooperation among

DOI: 10.54254/2755-2721/2026.GU34000

technologists and business practitioners to make better use of artificial intelligence in customer analytics.

References

[1] Reichheld, F. F., & Sasser, W. E. (1990). Zero defections: Quality comes to services. Harvard Business Review, 68(5), 105-111.
[2] Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications, 36(2), 2592-2602.
[3] Breiman, L. (2001). Random Forests. Machine Learning, 45 (1), 5-32.
[4] LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep Learning. Nature, 521(7553), 436-444.
[5] Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H. (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. Journal of Marketing Research, 43(2), 204-211.
[6] Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and Regression Trees. Wadsworth.
[7] Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory, 55, 1-9.
[8] Tsai, C. F., & Lu, Y. H. (2009). Customer churn prediction by hybrid neural networks. Expert Systems with Applications, 36(10), 12547-12553.
[9] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929-1958.
[10] Liu, Y., et al. (2022). A fused attentional deep learning model (AttnBLSTM-CNN) for customer churn prediction. IEEE Access.
[11] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.
[12] Jajam, A. V., et al. (2023). An ensemble model integrating stacked bidirectional LSTMs and RNNs with arithmetic optimization for churn prediction. Expert Systems with Applications.
[13] Zhao, Y., et al. (2023). Multi-head self-attention nested LSTM for telecom customer segmentation and churn prediction. IEEE Transactions on Neural Networks and Learning Systems.
[14] Van-Hieu, V., et al. (2024). Deep learning ensemble for customer churn prediction in banking. Data & Knowledge Engineering.
[15] Arrieta, A. B., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115.
[16] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765-4774.
[17] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144).
[18] Dwork, C., & Roth, A. (2014). The Algorithmic Foundations of Differential Privacy. Foundations and Trends in Theoretical Computer Science.
[19] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (pp. 1273-1282).
[20] Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359.