# Advancements of Customer Churn in the Telecommunications and Financial Industries Based on Machine Learning

Yankai Wang[a]
*Department of Audit, Zhejiang University of Finance and Economics, Hangzhou, China*

Keywords: Customer Churn Analysis, Machine Learning, Predictive Model.

Abstract: Faced with increasingly fierce market competition, customers often frequently have a variety of options when selecting items and services. The issue of customer churn has become a pressing concern for the majority of businesses, as seen by financial organizations (such as banks) and telecommunications companies. This paper provides an overview of the application of machine learning techniques in predicting customer churn in the telecommunications and financial industries. The purpose is to summarize the most advanced methods and evaluate their effectiveness in predicting customer turnover. In the telecommunications field, literature emphasizes the application of K-means clustering in customer segmentation, followed by predictive models such as XGBoost and Adaboost, which have been shown to perform well in capturing complex relationships in customer data. Similarly, in the financial field, random forests, support vector machines (SVM), and LightGBM are widely popular for their ability to handle large-scale datasets and nonlinear patterns, thereby improving the accuracy of customer churn prediction. Based on existing research, this paper discusses the challenges and improvement methods of artificial intelligence and machine learning in the field of customer churn prediction and analysis.



## 1 INTRODUCTION

Customers are an important resource of a company, and it is the key to the sustainable operation of enterprises, which can bring a large number of profits to the company. Customer churn is caused by the implementation of various marketing methods by the enterprise, which leads to the termination of cooperation between customers and the enterprise. This may be because they are not satisfied with the services or goods received, or because they have received more satisfactory substitutes from other enterprises. In the context of the digital information age, people are increasingly exposed to resources and have access to information, and there is a phenomenon of customer loss in various industries. Because customer churn not only means the company needs to incur new acquisition costs, but also spends more costs to recover customers. So, in the face of increasingly fierce competition in today's market, leaders of various enterprises are paying an increasing number of the attention to the issue of customer churn. Therefore, studying the characteristics of lost customers, analyzing their reasons for loss, and establishing appropriate and effective predictive models have gradually become an important topic in the field of business analysis.

Benefit by the popularity of artificial intelligence technology, machine learning algorithms are increasingly being applied in all works of life. Machine learning is a technique to explore how computers detect current knowledge, gain new knowledge, continually improve performance, and achieve self-improvement. It employs computers to replicate human learning activities (Chen, 2007). For example, Random Forests (RF), Logistic Regression (LR), K-Nearest Neighbor (KNN), Decision Trees (DT) are commonly used techniques in machine learning, and these algorithms are applied in many aspects. For instance, in the area of smart healthcare, Zheng et al. proposed a new method for testing Alzheimer's disease based on the GSplit LBI algorithm (Zheng, 2020). In the financial area, Manas et al. used KNN, Support Vector Machine(SVM), DT, and RF to predict bank customer churn, providing new ideas and methods for user churn

[a] https://orcid.org/0009-0005-0992-3295



Wang, Y.
Advancements of Customer Churn in the Telecommunications and Financial Industries Based on Machine Learning.
DOI: 10.5220/0013332000004568
In *Proceedings of the 1st International Conference on E-commerce and Artificial Intelligence (ECAI 2024)*, pages 611-616
ISBN: 978-989-758-726-9

(Rahman, 2020).Li et al. utilize five models–LR, RF, SVM, Least Absolute Shrinkage and Selection Operator (LASSO), and Light Gradient Boosting Machine (LGBM)–for machine learning in the field of electric vehicles in order to find pertinent characteristics that influence the sales of various manufacturers. They get the results by applying the voting procedure to the chosen features (Li, 2022). In the field of geology, Long et al. examined pre-earthquake ionospheric data and created a seismic ionospheric anomaly classification and prediction model based on gradient boosting decision tree algorithm (Long, 2022). Finally, Tsai et al. developed a customer churn prediction and reaction framework that consists of three stages: customer churn understanding, customer churn response, and customer churn prediction in the field of customer analysis. To increase customer service efficiency, this framework can be utilized to generate personalized or customized goods and services (Tsai, 2019).

The aim of this paper is to provide a comprehensive review in this field. The rest of the paper is arranged as follows: Section 2 outlines the methods used for customer churn prediction analysis. Section 3 compares various methods, describes their advantages and disadvantages, as well as some challenges or challenges faced by the field and future prospects. Finally, the conclusions of this work and future work are discussed in Section 4.

# 2 METHOD

## 2.1 Introduction of the Machine Learning Workflow

### 2.1.1 Data Collection

Data collection is the cornerstone of machine learning, which involves collecting, organizing, and preparing data for training and evaluating models. High quality, diverse, and representative datasets are the solid foundation for machine learning algorithm learning and optimization, which can improve the predictive accuracy and generalization ability of models.

### 2.1.2 Data Processing

Data processing is a key link in the fields of data science and artificial intelligence, which involves extracting, cleaning, transforming, and organizing data from raw data sources for subsequent data analysis and model training. Data cleaning and preprocessing are the two main stages of data processing, and they play an important role. Data cleaning includes removing noise, missing values, and errors from data, as well as organizing and standardizing data formats. Data preprocessing includes feature engineering, feature selection, normalization, and standardization operations on data to facilitate model training and analysis.

### 2.1.3 Model Building

Building a model is the process of generating a machine learning model from a set of feature vectors extracted from training data, which is used to predict test data. Firstly, it is necessary to determine what kind of model to establish, that is, to choose a suitable model. There are many machine learning models that can be classified from multiple perspectives.

(1) Learning process: Supervised Learning, Unsupervised Learning, Semi-supervised Learning.

(2) Task type: Clustering, Classification, Regression, Tagging.

(3) Model complexity: Linear Model, Non-linear Model.

(4) Model functionality: Generative Model, Discriminative Model.

### 2.1.4 Model Training

Model training is the process of training a model using a set of feature vectors generated by feature engineering. After multiple rounds of training with input feature vector sets, the internal parameters of the model gradually become fixed, and the model's response to the input also gradually stabilizes. Model training requires a considerable amount of time, mainly influenced by factors such as problem size, training conditions, and algorithm complexity.

### 2.1.5 Model Deployment

Model deployment refers to deploying a trained machine learning model to a production environment for practical use. Before the model is released, it needs to be exported from the training environment and then deployed to the production environment.

## 2.2 Customer Churn Prediction in Telecommunication

### 2.2.1 K-Means

K-Means is an unsupervised learning method that groups 'n' observations into k clusters, assigning each observation to the closest cluster center, or centroid, in an effort to minimize the variation within each cluster. Liu et al. employed 900,000 data items for various tasks such as feature extraction, feature selection, and data preparation. They suggested using K-means to cluster various customer groups, MIC and ratio to determine the ideal number of clusters, and factor analysis to determine which factors impact which consumer groups within that number of clusters (Liu, 2023).

#### 2.2.2 XGBOOST

Supervised machine learning techniques like XGBoost are commonly used for tackling classification, regression, and rank-based problems.It is a Gradient Boosting implementation using Decision Trees. The decision trees are used sequentially in this method (Sikri, 2024).

A hybrid architecture that has been proposed by Shimaa Ouf et al. may increase the precision of customer churn prediction analysis in the telecommunications. Effective data pretreatment approaches are applied in the construction of this framework, which combines the XGBOOST classifier with the mixed resampling method SMOTE-ENN. Two experiments are conducted using the suggested framework on three datasets from the telecom sector. In this study, classifier performance was investigated both before and after data balancing, introduced the impact of data balancing, determined which attributes are most important and influence customer turnover, and examined the speed-accuracy trade-off in hybrid classifiers (Ouf, 2024).

#### 2.2.3 AdaBoost

The AdaBoost algorithm combines many models that are not very powerful to form a very powerful model. During this process, AdaBoost pays special attention to data points that were previously misclassified, ensuring that these points receive more attention in subsequent training, thereby improving the overall learning performance.

Omid Soleiman garmabaki et al. investigated the elements that affect customer attrition in the telecom sector. They employed data mining classification techniques like support vector machines, K-nearest neighbors, and neural networks in this aim. Examine the outcomes using metrics like the ROC curve, accuracy, and precision. They further examined at how acceleration techniques, high-precision classifiers like neural networks, and data balancing interact with one another. Their most significant research contribution is the speed-accuracy trade-off approach they have developed for handling real-world hybrid classifier challenges. It evaluates the classifier's performance both before and after data balancing. They integrate the effective classifiers with the AdaBoost and XGBoost techniques after finding them. Based on every evaluation criterion, identify the combination that works best. According to the study's findings, performance can be greatly increased by utilizing a hybrid classifier combining AdaBoost and XGBoost(Soleiman-garmabaki, 2024).

### 2.3 Customer Churn in Financial Field

#### 2.3.1 Random Forest

It is a boosting algorithm that combines several weak classifiers to improve performance. It selects a random training sample subset to plant trees. Use parameter m to segment the nodes used for separating the total number of descriptors, where the selected separation features are much smaller than the total number of features. The standard random forest integrates multiple tree prediction factors that learn from the same distribution in the forest (Thomas, 2023).

de Lima Lemos et al. investigated customer churn prediction in the banking sector using a special customer level dataset from a major Brazilian bank. In order to make fair and reasonable comparisons between algorithms, they raced with a variety of supervised machine learning algorithms using identical evaluation and cross validation parameters. Research have shown that random forest technology performs better in a number of indicators than decision trees, logistic regression, k-nearest neighbors, elastic networks, and support vector machine models. A survey shows that customers who have closer relationships with banks have more resources, including goods and customer services. They are less likely to cancel their checking accounts and borrow more money from banks. Their model has a major economic impact, as it roughly estimates a potential loss of up to 10% on the operating performance recorded by Brazil's largest bank in 2019. The study's findings support the necessity of funding upselling and cross-selling initiatives that target present clients. These tactics might benefit client retention in the long run (de Lima Lemos, 2022).

#### 2.3.2 Support Vector Machine

Support Vector Machine is a binary classification model that aims to find a hyperplane to segment samples, with the principle of maximizing the interval. The goal of SVM is to find this hyperplane. SVM is very good at building hyperplanes or sets of hyperplanes in high-dimensional domains, which makes it useful for a variety of applications including regression and classification. Processing non-linear separable data by transferring it to a higher dimensional space where linear separation is possible is one of SVM's primary advantages.

Vikas Ranveer Singh Mahala et al. presented a thorough case study carried out at supermarkets, introducing a new type of golden membership and using sophisticated research and machine learning techniques to pinpoint possible clients and identify factors that influence customer reactions to new supermarkets. They developed a predictive model to measure the likelihood of customers responding positively (Singh Mahala, 2024).

#### 2.3.3 Light Gradient Boosting Machine (LightGBM)

LightGBM is an excellent tree based gradient boosting framework. Compared with existing boosting frameworks, the advantage of LightGBM lies not only in higher efficiency and accuracy, but also in lower memory consumption. In order to further improve the speed of the framework, people conducted learning experiments by setting specific parameters on multiple machines. The LightGBM running on this basis achieved linear acceleration (Changran J, 2022).

Ren et al. creatively integrated new supply chain data from suppliers and customers in businesses, adopting an integrated machine learning framework called LightGBM to build a predictive model for credit ratings using an algorithm. Utilizing data from North American listed firms between 2006 and 2020, they discovered that incorporating supply chain details from the year prior enhanced forecast accuracy when compared to incorporating supply chain details from that year. They discovered that models built using data from that year fared better in the wake of the COVID-19 epidemic, suggesting that the pandemic may have sped up the supply chain's diffusion of credit risk. Furthermore, studies have shown that when it comes to predicting target organizations' credit ratings, supplier information is more valuable than customer information (Ren, 2023).

## 3 DISCUSSIONS

### 3.1 Limitations and Challenges

#### 3.1.1 Interpretability

Interpretability is crucial in customer churn prediction analysis for understanding the reasons for predictions, model weaknesses, and repairing systems. The interpretability of algorithms refers to the ability of people to understand how algorithms make decisions. This is particularly important for disciplinary majors, as certain decisions may require complex reasoning processes. If the algorithm is not interpretable, professionals may not be able to understand and trust its results. Implementing highly interpretable algorithms is a complex task. Moreover, some machine learning algorithms are inherently black box models, making it difficult to explain their internal operational mechanisms. Secondly, some algorithms may have issues with local optima, which may result in their inability to provide accurate explanations in certain situations. For instance, multinational telecommunications companies deploy the same customer churn prediction model in two countries, but due to cultural differences, data biases, and compliance differences, the model that performs well domestically is not applicable abroad. So it's difficult for managers of overseas companies to trust this model when making predictions.

#### 3.1.2 Applicability

Applicability is directly related to the effectiveness and performance of machine learning algorithms in practical applications. Applicability refers to the ability of an artificial intelligence system or algorithm to effectively operate and produce expected results in a specific environment, task, or scenario. It involves the universality, flexibility, and stability of technology under different conditions. A home appliance company (mainly selling dishwashers) attempted to directly apply recommendation algorithms based on the US market to the Chinese market, but failed to consider cultural and consumer habits, resulting in poor applicability of the recommendation system, decreased user satisfaction, and sales performance that did not meet expectations.

#### 3.1.3 Privacy

Yang et al. proposed that in customer churn prediction analysis, researchers often directly use real data, which can easily lead to the leakage of user privacy data. Customer data usually collects users' basic attributes and behavioral data. Data involving user privacy needs to be protected to prevent leakage from causing losses and harm to customers (Yang, 2024). For example, when building a model, a sales company leaked customer names, phone numbers, and other information, causing customers' communication devices to be constantly harassed by advertisements and junk information. This can affect customer satisfaction with the company and products, increase customer churn, and ultimately lead to a decline in market competitiveness.

## 3.2 Future Prospects

### 3.2.1 Expert System, SHAP

To address the issue of poor interpretability in the above models, it is possible to optimize this part through expert systems or SHAP algorithms. Expert systems are among the earliest types of artificial intelligence and are widely used in a variety of sectors, including industry, healthcare, education, and finance (Duda, 1983). It can convey in-depth understanding of a complicated system and use an inference engine to produce the desired outcomes (Xiang, 2023).

Liu et al. proposed SHapley Additive exPlanations (SHAP) is a technique for illuminating machine learning models' predictions. Strong interpretability and independence from the predictive model are its strengths (Liu, 2024). Applications of SHAP include customer churn analysis, business analysis, and management. These applications have successfully enhanced the interpretability of machine learning and its capacity to discern the causal linkages between forecast results. In contrast to different techniques that rely on the internal structure of the model to evaluate feature importance, SHAP interpretability technology effectively eliminates interpretability differences that may arise due to various model designs by precisely estimating the marginal contribution of every input feature to the model prediction outcomes, offering a more consistent and all-encompassing method of evaluating feature relevance.

### 3.2.2 Transfer Learning, Domain Adaptation, Domain Generalization

In order to solve the issue of poor applicability of customer churn prediction analysis models, methods such as domain transfer, domain adaptation, and domain generalization can be used to improve the performance of the model.

**Transfer learning:** mainly focuses on how to learn new tasks on already trained models, thereby reducing the training time and data requirements of new tasks. The characteristics of transfer learning include:

(1) Reducing training data requirements: by training on the source domain dataset, transfer learning can reduce the training data requirements for new tasks.

(2) Reducing training time: by using already trained model parameters, transfer learning can reduce the training time for new tasks.

(3) Improving generalization ability: Transfer learning can achieve generalization in unseen domains, thereby enhancing the model's generalization ability.

**Domain adaptation:** refers to applying a model that has been trained in one domain (the source domain) to another domain (the target domain), despite the fact that the data distributions in these two domains are not the same. Optimizing the model for optimal performance in the target domain is the aim of domain adaptation.

**Domain generalization:** refers to the ability of a model learned on a task to be applied in unseen domains, thereby achieving cross domain knowledge transfer.

(1) Implementing cross domain knowledge transfer: Domain generalization can generalize in previously unseen domains, thereby achieving cross domain knowledge transfer.

(2) Improving model generalization ability: Domain generalization can improve the model's performance in unknown domains, which will strengthen its capacity for generalization.

### 3.2.3 Federated Learning

In order to address the risk of privacy breaches in machine learning federated learning can be introduced during the model building process to tackle this potential threat. Federated learning is a machine learning approach that makes it easier for several users to train together on the same model. Collaborative learning improves the training efficiency of models by reducing data volume, lowering communication costs, and increasing resource utilization. Its purpose is to promote collaboration and exchange among data parties in a distributed learning environment, thereby mutually enhancing the knowledge and experience of each party's data and forming a more cohesive whole. Overall, by merely exchanging model parameters or intermediate results, federated learning can achieve data privacy protection by building a global model based on virtual fusion data without the requirement to communicate with local individual or sample data.

# 4 CONCLUSIONS

This paper provides a comprehensive review of customer churn prediction analysis in the field of machine learning. The analysis of customer churn prediction using machine learning in the telecommunications and financial industries has shown promising results. In the telecommunications field, methods such as K-means clustering, XGBoost, and Adaboost have effectively identified churn patterns. Meanwhile, finance utilizes random forests, SVM, and LightGBM to improve prediction accuracy. Despite the success, challenges still exist, including data quality, model interpretability, and compliance with privacy regulations. Future directions include optimizing models for dynamic market conditions, enhancing model interpretability, and utilizing advanced artificial intelligence technologies such as deep learning for more detailed predictions. This constantly evolving pattern is expected to improve customer retention strategies and enhance business competitiveness.

# REFERENCES


Chen, K., & Zhu, Y. 2007. Overview of machine learning and related algorithms. Statistics and Information Forum, 05, 105-112.

Changran, J. 2022. Data analysis and machine learning in the context of customer churn prediction. In Proceedings of the 4th International Conference on Computing and Data Science (Part 3) (pp. 137-149). School of Naval Architecture, Ocean & Civil Engineering, Shanghai Jiao Tong University.

De Lima Lemos, R. A., Silva, T. C., & Tabak, B. M. 2022. Propension to customer churn in a financial institution: A machine learning approach. Neural Computing and Applications, 34(14), 11751-11768.

Duda, R. O., & Shortliffe, E. H. 1983. Expert systems research. Science, 220(4594), 261-268.

Li, Z. 2022. Research on sales strategy of electric vehicle target customers based on machine learning algorithm. Highlights in Science, Engineering and Technology, 22, 270-278.

Liu, Y., Dong, Y., Jiang, Z., & Chen, X. 2024. Interpretable predictive model for inclusions in electroslag remelting based on XGBoost and SHAP analysis. Metallurgical and Materials Transactions B, 55(3), 1428-1441.

Liu, Y., Fan, J., Zhang, J., Yin, X., & Song, Z. 2023. Research on telecom customer churn prediction based on ensemble learning. Journal of Intelligent Information Systems, 60(3), 759-775.

Long, Y., Zhang, Q., Dai, Z., & Rong, J. 2022. Investigation of ionospheric disturbance and seismic events based on machine learning. Highlights in Science, Engineering and Technology, 9, 37-42.

Ouf, S., Mahmoud, K. T., & Abdel-Fattah, M. A. 2024. A proposed hybrid framework to improve the accuracy of customer churn prediction in telecom industry. Journal of Big Data, 11(1), 70.

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.

Ren, L., Cong, S., Xue, X., & Gong, D. 2023. Credit rating prediction with supply chain information: A machine learning perspective. Annals of Operations Research, 1-30.

Sikri, A., Jameel, R., Idrees, S. M., & Kaur, H. 2024. Enhancing customer retention in telecom industry with machine learning driven churn prediction. Scientific Reports, 14(1), 13097.

Singh Mahala, V. R., Garg, N., Saxena, D., & Kumar, R. 2024. Unveiling marketing potential: Harnessing advanced analytics and machine learning for gold membership strategy optimization in a superstore. SN Computer Science, 5(4), 374.

Soleiman-garmabaki, O., & Rezvani, M. H. 2024. Ensemble classification using balanced data to predict customer churn: A case study on the telecom industry. Multimedia Tools and Applications, 83(15), 44799-44831.

Tang, T. 2023. Comparison of machine learning methods for estimating customer churn in the telecommunication industry. In Proceedings of the 5th International Conference on Computing and Data Science (Part 3) (pp. 201-206). College of Engineering and Applied Sciences, Stony Brook University.

Tsai, T. Y., Lin, C. T., & Prasad, M. 2019, November. An intelligent customer churn prediction and response framework. In 2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE) (pp. 928-935). IEEE.

Xiang, G., Wang, J., Han, X., Tang, S., & Hu, G. 2023. A novel optimization method for belief rule base expert system with activation rate. Scientific Reports, 13(1), 584.

Yang, B., Wang, Z., Cheng, Z., Zhao, H., Wang, X., Guan, Y., & Cheng, X. 2024. Customer churn prediction based on diffusion model generated data reconstruction. Computer Research and Development, 02, 324-337.

Zheng, W., Cui, B., Sun, Z., Li, X., Han, X., Yang, Y., ... & Alzheimer's Disease Neuroimaging Initiative. 2020. Application of generalized Split linearized Bregman iteration algorithm for Alzheimer's disease prediction. Aging (Albany NY), 12(7), 6206.