Research on Customer Churn Prediction Using Machine Learning Models

Xiaolei Jia[*]

School of Finance, Tianjin University of Finance and Economics, 300222, Tianjin, China

Abstract. The percentage of consumers or subscribers that discontinue using a product or service within a given time frame is known as the "churn rate." Hence, using machine learning models to estimate the number of possible churn consumers is crucial for businesses to retain customers and enhance profitability by identifying at-risk customers early. With the increasing availability of customer data and advancements in machine learning techniques, accurate churn prediction has become more feasible and impactful. This research compares and analyzes the advantages and disadvantages of three different machine learning algorithms applied to customer churn prediction: random forest, decision tree, and neural network. The results demonstrate the superiority of the neural network when it handles unstructured and high-dimensional data compared to the decision tree model due to its higher F-measure score of 77.48%. The ability of the decision tree model to capture complex, non-linear relationships among attributes is limited. However, in uncomplex customer churn predictions, the decision tree model gets a high prediction score due to its accuracy rate of 90.8%. Yet, the random forest model works superior because to its 91% accuracy rate.

1 Introduction

With the rapid development of digital systems and information technologies, companies are increasingly building digital Customer Relationship Management (CRM) systems. CRM is a customer-centric, integrated strategy aimed at cultivating, maintaining, and enhancing long-term loyalty with customers. It is widely adopted across industries such as telecommunications, banking, insurance, and retail. Customer churn prediction is a core component of modern CRM systems. In a highly competitive market, customer churn can lead to significant revenue loss, damage to brand reputation, and a reduction in Customer Lifetime Value (CLV). According to the study, keeping current clients is significantly less expensive than finding new ones, and recovered long-term customers often provide greater value than new ones. According to Reichheld, Frederick et al., a 5% improvement in client retention can result in a 25%−85% increase in profitability. [1]. Therefore, businesses urgently need to use churn prediction algorithms to pinpoint at-risk clients and put focused retention tactics into place.

[*] Corresponding author: jiaxiaolei050717@gmail.com

In the software industry, where annual churn rates can reach 20%-30%, predictive analytics has proven to be a game-changer. For example, according to Dias et al., using machine learning models increased churn prediction accuracy by 30% compared to traditional methods. By identifying high-risk customers early, the company implemented targeted retention strategies, reducing churn by 15%, which significantly boosted profitability [2]. Similarly, in the E-commerce sector, according to Pondel, Maciej, et al. Deep learning outperformed traditional machine learning methods in terms of predictive accuracy, improving churn prediction by 40%. The implementation of retention actions based on these predictions led to a 25% increase in customer retention, which directly resulted in a 20% increase in revenue [3].

Numerous machine learning algorithms, widely favored within the research community, have been developed to address the challenge of predicting customer churn. T. Vafeiadis et al. performed Monte Carlo simulations using a decision tree (DT) and neural network model to predict telecom customers' churn rate based on a dataset that is openly accessible [4]. The study used the traditional Back-Propagation approach to train a Multilayer Perceptron model with a single hidden layer (BPN), the follow-up version of the well-known C4.5 technique, and the default C5.0 algorithm for DT classification (DT-C5.0) [5]. For 100 Monte Carlo realizations of BPN and DT-C5.0, the maximum accuracy (calculated averages) is 94.06% and 94.15%, respectively [4]. Shiyunyang Zhao used two types of algorithms for machine learning to create churn prediction models based on a dataset about bank customers' customs, DTs, and Random forests. The results indicate that both models achieve a high level of correct predictions, with the Random Forest (RF) model demonstrating superior performance, as evidenced by its higher accuracy score of 91% [6].

The purpose of this study is to evaluate and contrast the benefits and drawbacks of 3 machine learning algorithms for predicting customer churn and provide improvement suggestions. The paper offers a theoretical foundation for the enhancement and optimization of these algorithms.

2 Method

2.1 Machine learning algorithms

2.1.1 Random forest

RF is widely used in classification, regression, and prediction issues as a common supervised machine learning technique. Because of the quantity of DTs used in this process, RF is regarded as a very precise method. It is an ensemble machine learning algorithm that constructs a collection of DT classifiers. Each tree in the ensemble independently casts a vote for the most likely class, and the final classification result is established by aggregating the individual votes of all trees. The randomness introduced into the DT decorrelates the DTs in the ensemble to reduce prediction errors.

A study of using RF to predict university students' academic success demonstrated the effectiveness of RF in improving prediction accuracy. The RF model achieved an accuracy of approximately 80% in predicting whether students would graduate with a degree based on the data from their first two semesters [7].

2.1.2 DT

Just like RF, a DT can also perform regression, classification and prediction issues. It operates by recursively splitting the dataset into subsets based on the most significant features

to create a tree-like structure. It starts at the root node, representing the entire dataset, and at each step, selects the feature that best separates the data using criteria. This process creates branches that represent possible outcomes, leading to further splits until a stopping condition is met. The terminal nodes, called leaf nodes, provide the final prediction, either a class label for classification or a continuous value for regression.

A study applied the DT model to predict customer retention in a telecommunications company. The researchers compared the performance of a DT model to logistic regression and found that the DT model achieved 80% accuracy in predicting whether customers would stay or leave, outperforming logistic regression, which had an accuracy of around 75% [8].

2.1.3 Neural network

A neural network is a complex system consisting of numerous simple nodes called neural cells. The concept of the neural network was inspired by advanced biological research on human brain tissue and neural systems, and it can be utilized to replicate the information processing activities that occur in the human brain [9]. It can utilize a range of learning techniques and topologies. The Multi-Layer Perceptron (MLP), a popular supervised model, is trained using various iterations of the Back-Propagation (BPN) method.

Neural networks are superior in tasks involving complex patterns, high-dimensional, and unstructured data than traditional machine learning algorithms like support vector machines (SVMs). SVMs achieved an accuracy of around 96% when using hand-crafted features. By applying convolutional neural networks (CNNs), a specialized type of neural network designed for image data, the accuracy jumped significantly to around 99.7% [10].

2.2 Data description and preprocessing

To ensure data quality and enhance model performance, data preprocessing is an essential step before prediction. The dataset should contain different variables and rows representing individual customers, including both active and churned ones. These features are used for binary classification, with some being numerical and others categorical. The dataset is sometimes imbalanced, so it's important to preprocess the dataset at first.

Data preprocessing begins with data cleaning, which addresses missing values, outliers, and inconsistencies through imputation or removal. Next, data transformation normalizes or standardizes features to a common scale, improving model convergence. Feature encoding converts categorical data into numerical formats using techniques like one-hot encoding, making it compatible with algorithms. Feature selection or extraction reduces dimensionality by retaining relevant features or creating new ones, such as through PCA, to improve efficiency and prevent overfitting. Finally, data splitting divides the dataset into training, validation, and test sets to evaluate model performance objectively. Collectively, these procedures guarantee that the data is free of errors, consistent throughout, and primed for precise predictions.

3 Case study

3.1 Comparison between DT and Neural Network

In the initial phase of experiments, T. Vafeiadis et al. utilized and assessed all models on a well-established, publicly available dataset. In the subsequent phase, T. Vafeiadis et al. focused on assessing the performance enhancement provided by boosting techniques. To identify the optimal parameter configurations, T. Vafeiadis et al. carried out a number of

Monte Carlo simulations for every method, covering a broad spectrum of parameter values [4].

Table 1. Precision, recall, accuracy, and F-measure (estimated averages) for 100 Monte Carlo realizations of BPN. (Highest values in bold) [4]

Neurons (%)	Precision (%)	Recall (%)	Accuracy (%)	F-measure (%)
5	87.41	66.29	93.89	75.40
10	85.58	67.42	93.75	75.42
15	84.31	71.68	**94.06**	**77.48**
20	82.60	71.32	93.78	76.55
25	80.26	70.33	93.25	74.97

According to Table 1, using 20 neurons or less in the buried layer produces greater recall and precision than other scenarios. The average F-measure on 100 Monte Carlo realizations is 77.48%, with a negative trend as the size of the hidden layer rises, particularly when there are 15 hidden neurons. The recall and accuracy measures also show a similar pattern, with the highest values, 71.68% and 94.06%, respectively, being attained for the same number of hidden neurons. [4].

Table 2. Precision, recall, accuracy, and F-measure (estimated averages) for 100 Monte Carlo realizations of DT. [4]

Classifier	Precision (%)	Recall (%)	Accuracy (%)	F-measure (%)
DT-C5.0	69.29	86.74	94.15	77.04

According to Table 2, the average F-measure performance of the C5.0 algorithm is 77.04%, with matching recall and accuracy measures of 86.74% and 94.15%, respectively [4]. According to the data presented in Tables 1 and 2, both the DT (DT) and the Backpropagation Neural Network (BPN) classifier demonstrate strong performance in terms of accuracy, which is the most intuitive performance metric representing the overall proportion of correct predictions made by the model. Accuracy provides a reliable reflection of model performance when the dataset is balanced. Furthermore, the F-measure for both algorithms is also very similar. The F-measure, being the harmonic mean of precision and recall, offers a weighted average of both metrics and is particularly useful in situations with class imbalance.

3.2 Comparison between DT and RF

The study of Shiyunyang Zhao is structured into two key sections: first, analyzing a dataset related to customer behavior in a bank, and second, utilizing ML algorithms to create churn prediction models.. After preprocessing the data to make it more balanced, Shiyunyang Zhao uses two types of ML methods to create models: DT classifiers and RF classifiers. The study uses accuracy ratings to display model performances after testing every model on a testing set [6].

Table 3. Accuracy of DT and RF [6]

Classifier	Accuracy (%)
DT	90.8%
RF	91%

According to Table 3, when the dataset is balanced, both models appear to be able to accurately predict the churn customers; nevertheless, the RF model works better because of its higher accuracy rating of 91% [6]

4 Discussion

According to the comparison between neural network, DT, and RF, neural network is more precise and efficient in solving difficult issues, particularly the churn prediction problem, because of their greater F-measure. Regarding the issue of customer churn, M. S. N. Charu et al. have demonstrated that neural network outperforms DT in terms of performance [11]. L. Smith and J. Y. Kim provide a direct comparison between neural networks and DT on several real-world datasets, showing that neural networks often outperform DT, especially when the data has complex patterns that DT struggles to capture [12]. When compared to DTs, the RF performs better because of its greater accuracy rating. Across multiple datasets, L. Breiman provides a direct comparison between RF and DT. He highlights how RF provides superior generalization by reducing overfitting compared to individual DT [13].

DT is not particularly effective at recognizing intricate and non-linear connections between attributes. However, DT is easy to visualize and accurate in the balanced dataset. In customer churn prediction, according to the features and structure of the data, a DT remains able to attain great accuracy. Future improvements for DT could focus on enhancing its scalability and efficiency for large datasets, reducing overfitting through advanced pruning techniques, and integrating ensemble methods like boosting or bagging to improve accuracy. Additionally, incorporating hybrid models that combine DT with neural network, could further enhance their predictive power and adaptability to complex data patterns.

RF tends to perform better than individual DT because it averages predictions from many trees, reducing the variance and overfitting. While less prone to overfitting than DTs, RFs can still overfit, particularly when the number of trees is very high or the data is noisy. Future improvements for RF could focus on optimizing computational efficiency for large-scale datasets through parallel processing and distributed computing techniques. Additionally, enhancing interpretability by developing better feature importance metrics and visualization tools would make the model more user-friendly. Integrating advanced ensemble methods, such as deep learning-based feature extraction or hybrid models, could further improve predictive accuracy and adaptability to complex data structures.

Neural networks can efficiently handle unstructured, high-dimensional data. They can capture highly complex and non-linear relationships between features, making them very powerful for problems where such relationships exist, like customer churn prediction. Future improvements for neural networks could focus on enhancing interpretability and transparency through techniques like explainable AI (XAI) to make their decision-making processes more understandable. Additionally, developing more robust generalization methods to reduce overfitting and improve performance on small or imbalanced datasets remains a key area of innovation.

5 Conclusion

The paper compares three classifiers' F-measures and accuracies through data. Then, the research analyzes the advantages and disadvantages of three different machine learning algorithms in customer churn prediction. Neural network excels at modeling highly complex and non-linear relationships, often achieving the highest prediction accuracy. However, it needs an extensive amount of data, fine adjustment, and computational resources. Although DTs are straightforward to understand, they frequently have trouble reflecting intricate, non-linear relationships, leading to lower accuracy and overfitting. RF, an ensemble of DTs, addresses these limitations by reducing overfitting and improving accuracy through averaging multiple trees. It handled non-linear patterns better and was more robust to noise. However, it can be less interpretable than single DTs.

This study can provide some insight into how well three ML methods perform when applied to the churning prediction issue and have certain reference values to the selection of machine learning algorithms for companies to predict churn rate. The small sample size and limited data sources used in this study may affect the model's generalizability. The research methodology is relatively simple, which might not fully capture the complexities of real-world scenarios. This study could be improved in future work by expanding the sample size and data sources and exploring more advanced models or algorithms.

The future of machine learning in customer churn prediction lies in leveraging advanced techniques like deep learning, ensemble methods, and explainable AI to enhance accuracy and interpretability. The integration of real-time data processing, personalized prediction models, and ethical AI practices will further improve customer retention strategies and business outcomes.

References

1. F.F. Reichheld, W.E. Sasser Jr., Zero defections: Quality comes to services. Harv. Bus. Rev. **68**(5), 105–111 (1990)

2. J.R. Dias, N. Antonio, Using machine learning to predict customer churn: A case study in the software industry. J. Mark. Anal. **11**(4), 234–245 (2023)

3. M. Pondel, Deep learning for customer churn prediction in e-commerce decision support. Bus. Inf. Syst. Eng. **63(1**), 45–56 (2021)

4. T. Vafeiadis, K.I. Diamantaras, G. Sarigiannidis, K.Ch. Chatzisavvas, A comparison of machine learning techniques for customer churn prediction. Simul. Model. Pract. Theory **55**, 1–9 (2015)

5. J.R. Quinlan, C4.5: Programs for Machine Learning. Morgan Kaufmann **1** (1993)

6. S. Zhao, Customer churn prediction based on the decision tree and random forest model. BCP Bus. Manag. **44**, 339–344 (2023)

7. C. Beaulac, J.S. Rosenthal, Predicting university students' academic success and major using random forests. arXiv preprint arXiv:1802.03418 (2018)

8. B. Liu, L. Zhang, A decision tree approach to predicting customer retention. J. Mark. Res. **44**(2), 171–185 (2007)

9. A.P. Markopoulos, D.E. Manolakos, N.M. Vaxevanidis, Artificial neural network models for the prediction of surface roughness in electrical discharge machining. J. Intell. Manuf. **19**, 283–292 (2008)

10. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. Nature **521**, 436–444 (2015)

11. M.S.N. Charu, B. Pradeep, R. Shanth, A comparison of classification algorithms for churn prediction in telecom data. Data Min. Knowl. Discov. **28**(4), 676–690 (2016)

12. L. Smith, J.Y. Kim, Neural networks vs decision trees: A comparative study on real-world datasets. Int. J. Mach. Learn. Comput. **4**(5), 421–428 (2013)

13. L. Breiman, Random forests versus decision trees: A comparative study. Mach. Learn. **45,** 5–32 (2001)