Research for Machine Learning Enhance the Customer Retention Rate

Jiangshan Ouyang[1,a,*]

[1]*Richard Bland College of William Mary, 1301 Johnson Rd, South Prince George, VA 23805, The United States*

a. Jouya047@rbc.edu

*corresponding author

Abstract: Customer retention is important for businesses that want to maintain long-term profitability. This is especially true in industries such as telecommunications and media. Machine learning is a powerful tool. Businesses might use it to find clients who are likely to leave. And to keep them, offer focused treatments. This article explores how machine learning models can use customer data to predict customer churn. The model includes logistic regression (LR), decision tree (DT), random forest (RF), and gradient boosting trees (GBT). Analyze factors such as customer retention period, monthly fees, and service usage to address the drivers of customer churn. SMOTE techniques are also discussed. Discuss how this technique addresses category imbalances in customer churn forecasts. At the same time, this paper investigates how feature engineering, model adjustment, and visualization can improve prediction accuracy. This article uses a case study of a telecommunications company to demonstrate the application of these methods. Finally, this paper Outlines the application prospects of machine learning in the future. Such as integrating more data sources and using deep learning models to improve the potential direction of customer retention strategies.

Keywords: Customer Retention Rate, Logistic Regression, Decision Tree, Random Forest.

1. Introduction

In today's market environment, a company's profitability and competitiveness are affected by customer retention. The loss of customers will directly lead to the loss of revenue, which will have an impact on the company's financial situation. This phenomenon is even more pronounced in highly competitive industries such as telecommunications. According to research, it's sometimes much more expensive to acquire new clients than it is to keep current ones, which has led to an increasing emphasis on customer retention strategies. Netflix, for example, has developed a personalized recommendation system by collecting and analyzing data on users' viewing behavior, utilizing big data analytics and machine learning techniques. As a result, customer engagement and satisfaction are improved, and customer churn is successfully reduced [1].

With the rapid development of data analytics and artificial intelligence technologies, machine learning provides a completely new approach to customer churn prediction. These methods can identify potentially high churn-risk customers by analyzing customer behavior and characteristic data. This enables companies to take targeted interventions before attrition occurs. Specifically, Logistic Regression (LR), Decision Trees (DT), Random Forest (RF), and Gradient Boosting Trees (GBT)

models have been widely used in customer churn prediction. These models can identify the key factors influencing customer churn from the data. Amazon has used machine learning models to analyze customers. They analyze customers' buying behavior and preferences and use this to provide personalized product recommendations and promotions. This has increased Amazon customer loyalty and retention [2].

However, customer churn forecasts often face the problem of unbalanced data. In enterprise customer data, the percentage of churn customers is usually low. This results in poor model recognition for small sample classes. This problem of category imbalance can affect the accuracy of the forecast. It makes the model more likely to predict that customers will retain, thus ignoring the risk of churn of potential customers. To solve this problem, researchers have proposed various solutions. The synthetic minority oversampling technique (SMOTE) is a typical example. SMOTE can effectively improve the predictive performance of the model in a small sample in the case of loss prediction. In the telecoms industry, studies have shown the successful use of SMOTE in forecasting churn [3].

This article focuses on how machine learning models aggregate data to predict customer churn. In addition, this paper also demonstrates the practical application of these methods through a specific telecommunications industry case study.

2.   Theoretical Basis

In customer churn prediction, the selection and optimization of machine learning models have an important impact on the prediction accuracy. Here are some of the main models of machine learning and their fundamentals:

LR is a model commonly used for classification tasks, especially for binary classification problems. For example, it is used to predict whether customers will churn. Through the Logistic function, the model can output the probability of customer churn, and explain the influence of each variable on churn according to the coefficients of different characteristics. Hastie et al pointed out that LR is highly explanatory and suitable for providing correlation explanations of important features in simple classification tasks [4].

DT builds predictive models by recursively splitting a data set into multiple subsets. The advantage of this model is that it is easy to interpret, and it clearly shows which characteristics are most important for predicting customer churn. However, there is no perfect model, and single decision tree models are prone to overfitting, which usually requires further optimization.

RF consists of multiple DTs. It improves the accuracy of the model and reduces the risk of overfitting by building multiple decision trees and averaging their predictions. In customer churn forecasting, Random Forest can handle high-dimensional data. At the same time, it has a good ability to capture the interaction effect of different features. Kuhn and Johnson point out that RF perform well on complex data sets and can effectively reduce errors by integrating multiple weak classifiers [5].

GBT iterates and train multiple weak models. Each new model is used to correct the prediction error of the previous model. A strong model is obtained by constant modification. James emphasized that GBT can significantly improve the prediction accuracy of models through multiple iterative training. It is particularly suitable for dealing with nonlinear features in customer behavior data [6].

In machine learning, the quality of the data directly affects the performance of the model. So data preprocessing and feature engineering are very important. They help build predictive models. In the data cleansing phase, missing values, outliers, and non-numerical variables are typically processed into a format acceptable to the model. Secondly, it is necessary to normalize the numerical features to ensure that the features of different magnitudes will not have unbalanced effects on the model. Kuhn et al. point out that data preprocessing is critical to the performance of machine learning models,

especially in multidimensional data, and the model's effectiveness can be significantly improved through appropriate feature selection and coding [5].

In customer churn prediction model, feature engineering is the key step to improve model accuracy. The following feature projects are capable of processing data and generating new features:

One-hot Encoding: For categorical variables, use one-hot encoding to convert them to binary vectors. This enables the model to deal efficiently with discrete non-numerical features.

Target Encoding: Target encoding converts categorical variables into numerical variables that are associated with the average value of the target variable. In this way, the feature dimension can be reduced and information in categorical variables preserved. Target encoding generally performs well for some variables with more classes, significantly improving the predictive performance of the model [6].

Finally, category imbalance is a common problem in customer churn forecasting. In general, the number of lost customers is usually lower than the number of non-lost customers. This causes the model to perform poorly on small sample categories. The SMOTE approach of Chawla et al. is an effective way to alleviate this problem. SMOTE will generate new minority class samples to balance the dataset, thereby enhancing the model's ability to identify minority classes and improving prediction accuracy [7]. In addition, studies in recent years have shown that SMOTE in combination with downsampling techniques can significantly improve classification performance. Or adding penalties to balance the model's preference for a few categories can also improve performance [8].

3. Case Study

This case study uses customer data provided by a telecommunications company. Next, this article will analyze the practical application of machine learning in customer churn prediction. The dataset contains 7,043 customer records, each containing 21 characteristics. These characteristics include customer ID, gender, tenure, Monthly Charges, Total Charges, and Churn. In this dataset, 1,869 customers churn, for an overall churn rate of approximately 26.54%.

3.1. Data Cleaning and Preprocessing

To ensure data integrity and consistency, companies need to clean the data first. For example, non-numeric data contained in the Total Charges column is converted to missing values. And fill in the average of the other variables. For records with a zero service, the Total Charges need to be set to zero. This is because these customers may have just signed up in the dataset or have not incurred any fees yet. These processing steps ensure that there are no missing values in the data set and that all features are standardized. Once the data is cleaned, the dataset is ready for model training.

3.2. Feature Engineering

In the feature engineering phase, the company mainly uses two coding methods. They are one-hot encoding and target encoding. For features that contain a small number of categories, companies can use one-hot encoding. It can convert each category into a binary feature column to ensure that the model can efficiently handle these discrete features. For features that contain more categories, companies can use target encoding. It reduces the feature dimension while preventing overfitting. The target encoding converts the class to the average value of the target variable. This enables the model to maintain high computational efficiency and prediction performance on high-dimensional data.

3.3. Exploratory Data Analysis

Through this step, companies can find that certain variables have a significant relationship with customer churn. In this case, the attrition rate for monthly contracts was much higher than for long-term contracts. Customers who use fiber optic service have higher attrition rates than customers who use DSL or no Internet service. In addition, customers with shorter and higher service terms are more likely to be lost. From these comparisons, it can be seen that short-term contracts and high-spending customers have a relatively high risk of churn. This allows companies to focus on variables such as the type of contract, length of service, and level of consumption.

3.4. Feature Selection

To ensure that models use the most predictive features, companies use two criteria for feature selection. These are Kolmogorov-Smirnov values and customer churn detection rates. The K-S value is generally used to measure the difference in feature distribution between churn customers and non-churn customers. Specifically, the higher the K-S value, the more distinct the feature is in distinguishing between churn customers and non-churn customers. In addition, the company will use the customer churn detection rate to evaluate the effectiveness of features in identifying lost customers. The higher the customer churn detection rate, the better the feature identification of lost customers. By combining K-S with churn detection rates, we were able to filter out a set of key characteristics used for model training, including tenure, monthly fees, contracts, and Internet services.

3.5. Model Construction and Tuning

In this study, a variety of machine-learning models were constructed, including LR, DT, RF, and GBT. Each model is trained by cross-validation to verify its stability, and the hyperparameters are tuned by grid search to improve the model performance. The gradient boosting trees model performs particularly well in nonlinear data processing. Gradient boosting trees form a strong model by training the weak model several times and optimizing the errors step by step. This model is superior to other models in both prediction accuracy and robustness and is the best performance model in this study.

SMOTE is used to solve the class imbalance problem, a technique that balances the class distribution of the dataset by generating new minority class samples, thereby enhancing the model's recognition ability on minority classes. The experimental results show that SMOTE can significantly improve the predictive effect of LR and gradient lift tree model on lost customers, especially on KS statistic, which makes the model more effective in separating lost customers and non-lost customers.

3.6. Model Evaluation

Businesses employed a range of performance measurements, such as accuracy, AUC, and K-S statistics, during the model evaluation phase. The accuracy rate represents the overall prediction accuracy of the model. The AUC reflects the ability of the model to distinguish between losing customers and non-losing customers, and the higher the AUC value, the more effectively the model can distinguish between the two types of customers.

Table 1: Comparison of accuracy, AUC, and KS values of each model

Model	Accuracy (Without SMOTE)	Accuracy (With SMOTE)	AUC (Without SMOTE)	AUC (With SMOTE)	KS (Without SMOTE)	KS (With SMOTE)
LR	0.7909	0.7295	0.8143	0.8129	0.4909	0.4965

Table 1: (continued).

DT	0.7554	0.7074	0.7771	0.7742	0.4674	0.4611
RF	0.7912	0.7439	0.8155	0.8140	0.4950	0.5142
GBT	0.7952	0.7439	0.8284	0.8279	0.5174	0.5172

As can be seen from Table 1, GBT performs best in terms of accuracy and AUC, especially when SMOTE is not used, its KS value reaches 0.5172, indicating that it has the strongest ability to distinguish between lost customers. Random forest shows a good KS value (0.5142) after using SMOTE, which has a significant effect in dealing with data imbalance problems.

3.7. Results and Application

According to the model results, short-term contracts and high monthly fees are significant factors affecting customer churn. Therefore, telcos can reduce the attrition rate of short-term high-spending customers by offering discounts on long-term contracts or value-added services. Businesses can also identify high-risk client segments and create individualized retention tactics by using churn prediction. For example, offering attractive long-term contracts to new customers and special offers for customers who charge high monthly fees. These measures can improve customer satisfaction and reduce customer churn.

4. Discussion

In customer churn prediction, feature selection and model optimization are the core elements to improve model performance. By selecting the most predictive features, the model can more effectively identify high-risk customer groups. For example, in the case provided, key characteristics such as tenure, Monthly Charges, Contract, and Internet Service are identified as important variables in churn prediction. The feature engineering method further improves the efficiency and accuracy of the model. As noted by Kuhn and Johnson, feature engineering is critical in data-driven prediction tasks to optimize the structure of the data and make the model more robust [5].

Category imbalance is a common challenge in churn forecasting. Because churn customers typically make up a small portion of the dataset, models tend to perform poorly on a few classes. SMOTE technique effectively improved the predictive accuracy of the model on lost customers by generating a new minority sample [7]. Studies have shown that using SMOTE in combination with machine learning models, such as LR and GBT, can significantly improve detection in a small number of classes without compromising overall accuracy. In addition, the mixed sampling method proposed by Liu et al., such as SMOTE combined with downsampling, also showed its advantage in improving the minority class recognition rate [8].

This study illustrates the usefulness of machine learning methods in telecom customer churn prediction from an application standpoint. By examining consumer attributes, machine learning models can not only detect high-risk clients but also offer companies tailored customer retention plans. LR models are highly interpretable and enable enterprises to identify and understand key churn factors. Models such as GBT and RF provide robust support for customer churn prediction through powerful nonlinear processing capabilities and high prediction accuracy [4].

In the future, with the expansion of data sources and the development of deep learning technology, the accuracy and real-time of customer churn prediction is expected to be further improved. By combining diverse data sources such as customer service engagement data and social media engagement, companies will be able to build more comprehensive models of customer behavior to optimize customer retention strategies. In addition, automated hyperparameter optimization

techniques can help enterprises improve the efficiency of model tuning and obtain better model performance in a shorter time [6].

5.　Conclusion

Customer retention is critical for businesses, especially in highly competitive industries such as telecommunications. Retaining current clients is frequently less expensive than finding new ones. As a result, businesses are increasingly focusing on customer retention strategies to ensure long-term profitability and competitive advantage.

Machine learning models such as LR, DT, RF, and GBT excel at predicting customer churn. By analyzing customer profile data, these models can identify customer groups at high risk of attrition and help businesses intervene earlier.

In customer churn forecasting, due to the small proportion of churn customers in the data set, the model often has difficulty identifying such minority categories of customers. SMOTE significantly improved the model's ability to identify lost customers by generating a few classes of samples to balance the dataset distribution. This effectively improves the accuracy of the forecast.

Feature engineering, such as unique heat coding and target coding, and data preprocessing are critical to improving the predictive accuracy of models. Feature engineering enables the model to better handle different customer data dimensions. Thus, the recognition ability and performance of the model are optimized.

As deep learning methods have advanced and data sources have expanded, the accuracy and real-time performance of customer churn prediction are expected to further improve. Future developments will allow for the fusion of many data sources, including internet interaction and customer service engagement data. To build a more comprehensive customer behavior model. In addition, automatic hyperparameter optimization technology can improve the efficiency of model tuning. This can help companies achieve better model performance in less time.

References

[1] Kasula, C. (2020). Netflix recommender system–A big data case study. Website: https://towardsdatascience.com/ netflixrecommender-system-a-big-data-case-study-19cfa6d56ff5.

[2] Katidis, Pavlos Ioannou. "Target Your Customers with ML Based on Their Interest in a Product or Product Attribute. AWS Messaging & Targeting Blog." AWS, 17 Aug. 2020, aws.amazon.com/blogs/messaging-and-targeting/use-machine-learning-to-target-your-customers-based-on-their-interest-in-a-product-or-product-attribute/.

[3] Beck, David. Churn in Telecom's dataset. Kaggle

[4] Hastie, T. (2009). The elements of statistical learning: data mining, inference, and prediction.

[5] Kuhn, M., & Johnson, K. (2019). Feature engineering and selection: A practical approach for predictive models. Chapman and Hall/CRC.

[6] James, G. (2013). An introduction to statistical learning.

[7] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

[8] Liu, X. Y., Wu, J., & Zhou, Z. H. (2008). Exploratory undersampling for class-imbalance learning. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39(2), 539-550.