Applying Machine Learning Methods to Forecasting Customer Churn for a Telecommunications Company

Ruslan Bazhenov[1[0000-0003-2668-1142]], Nelly Sedova[2[0000-0003-4612-7843]], Viktor Sedov[3[0000-0002-1503-6626]], Evgeniy Lavrov[4[0000-0001-9117-5727]] and Andrey Dorofeev[5[0000-0002-8498-3301]]

[1] Sholom-Aleichem Priamursky State University, 70A, Shirokaya street, Birobidzhan, 679015, Russian Federation
r-i-bazhenov@yandex.ru
[2] Maritime State University named after G.I.Nevelskoy, 50A, Verkhneportovaya street, Vladivostok, 690059, Russian Federation
nellyfish81@mail.ru
[3] Maritime State University named after G.I.Nevelskoy, 50A, Verkhneportovaya street, Vladivostok, 690059, Russian Federation
sedov@msun.ru
[4] Sumy State University, 2, Rymskogo-Korsakova street, Sumy, 40007, Ukraine
prof_lavrov@mail.ru
[5] Irkutsk National Research Technical University, 83, Lermontova street, Irkutsk, 664074, Russian Federation
dorbaik2007@mail.ru

Abstract. The paper presents a brief overview of existing approaches to predicting customer churn using the example of a telecommunications company. The authors provide research on churn forecasting using 11 different machine learning methods. For training, data containing 20 different information parameters about clients were used. The quality of education was assessed using the traditional Area Under the Curve characteristic. The paper also provides research results confirming that the use of ensembles of machine learning methods increases the quality of predicting customer churn.

Keywords: Machine Learning, Customer Churn, Telecommunications Company, Area Under The Curve, An Ensemble of Techniques.

1 Introduction

The number of clients for any company is undoubtedly an important parameter, and the more clients, the higher the company's profit. There is no monopoly on the provision of mobile services to customers, high-speed Internet access or cable television, therefore, the so-called outflow of customers is possible for one reason or another, most often associated with the transition to another telecommunications company offering more favorable conditions from the point of view of customers ... To ensure the stable operation of a telecommunications company, an analysis of the customer base is necessary, for example, to ensure targeted promotions, as well as an analysis of the reasons for customer churn, including with the ability to predict the number of customers who have left and the reasons for switching to the services of another telecommunications company. According to experts, attracting one new client costs companies on average seven times more (in some cases, from 5 to 20 times more) than retaining an existing one [1-3]. Reducing customer churn by 5% increases the company's profits from 25% to 85% [4]. Therefore, understanding how exactly to maintain customer engagement is a natural foundation for developing strategies for customer retention.

In this regard, there is a number of scientific studies aimed both at predicting customer churn and predicting their preferences in any specific services. For example, in work [5], the authors try to understand the behavior of customers when paying bills, using such machine learning methods as logistic regression, one rule and support vector machines, they proposed an analytical approach to studying and predicting the payment behavior of customers. As input for the analytical system, they use: customer ID (a hashed unique number indicating each customer), action type (such as SMS, IVR, phone calls, service cut related and legal actions), the date of the action's execution, stage changer fl ag (such as payment through the banking system, unpaid invoice occurrence, or action timer fl ag). The feasibility of using machine learning methods is ensured by their ability to detect patterns in data of various natures and universes [6].

In [7], on the basis of Call Data Records data, various types of customers are determined, which subsequently affects the analysis of the possibility of their outflow and the impact on other customers of the company. Authors include these types of customers: follower, standard, leader, core and important customer. The authors note that customers who interact with the Leader and Important categories are more likely to churn after an influential member who also left the company. For clustering, the authors used neural networks such as linear perceptron, multilayer perceptron and networks with radial basis activation function. In the work of the authors [8], using the random forest classifier, decision tree classifier, gradient-boosted tree classifier and multiplexer perceptron, clients were segmented according to the "time-frequency-monetary" approach, while "Time" characterizes the total of calls duration and Internet sessions in a certain period of time, the "Frequency" parameter characterizes the frequency of using services frequently within a certain period, and the "Monetary" parameter is determined by the amount of money spent during a certain period.

It should be noted that there is often a situation where the assessment of the forecast of customer churn is carried out according to completely different criteria. For example, in [9], forecasting is based on the following information: region where the customer lives; time since the customer joined the operator (in months); average revenue; how long did the customer not pay the bills; amount that the customer is overdue; number of times the service was disconnected, etc. In the study [10], the following parameters are used to predict customer churn: state the US state, in which, the customer resides, the remaining seven-digit phone number, total number of calling minutes used during the day, the billed cost of daytime calls, total number of calling minutes used during the nighttime, and a few others. Moreover, some of these criteria may not be available to telcos, making it difficult to determine how, in practice, a telco will forecast these models. It is also worth mentioning several studies [11-12] on the use of customer analysis technologies in the banking sector, for example, to predict the likelihood of outflow based on information related to the socio-demographic parameters of customers, their activity in obtaining banking information, information about their salaries, and etc.

There is no unambiguous option for the machine learning methods used, there are a lot of them, it is almost impossible to choose an empirically suitable method, it is necessary to conduct training, in a practical way, selecting the optimal architecture and parameters of the machine learning method.

This paper presents the results of predicting the churn of customers of a telecommunications company using machine learning methods, since the latter are able to use statistical data, to determine dependencies between data of different nature.

2 Preparation of training sample

The problem of predicting customer churn can be formulated as a binary classification problem [9]. The solution to the problem is to classify customers with the corresponding characteristics (parameters) $x \in X$ to one of the two classes $Y = \{no\ churn, churn\}$.

To solve such a problem, in this work, we used a dataset consisting of 21 columns and 7043 rows made publicly available by IBM and containing information about customers of a telecommunications company (available at https://www.kaggle.com/blastchar/telco- customer-churn). Information about each customer includes the following characteristics, which are identifying signs for predicting churn (they are presented in Tables 1-3), and, of course, each customer is assigned a unique customer identifier (customerID).

Also, data on the period of use of the services of a telecommunications company were used as input information, the parameter is calculated in months (tenure); the size of the client's monthly fee (MonthlyCharges) and the final amount of payments for the entire period of work with the client (TotalCharges). For each set of input information, consisting of 20 indicators, there is a predetermined output - Churn - characterizing whether the client left the telecommunications company with these parameters or not.

Table 1. Description of identification signs that have two answer options.

No.	Identification sign	Churn customers	Non Churn customers
1	Gender (gender) of the client (gender)	Male – 50.7 % Female – 49.3 %	Male – 50.2 % Female– 49.8 %
2	Whether the user is a Senior Citizen	0 – 87.1 % 1 – 12.9 %	0 – 74.5 % 1 – 25.5 %
3	The client has a partner (Partner)	Yes – 52.8 % No – 47.2 %	Yes – 35.8 % No – 64.2 %
4	Client has Dependents	Yes – 34.3 % No – 65.7 %	Yes – 17.4 % No – 82.6 %
5	The indicator of the client's phone number (PhoneService)	Yes – 90.12 % No – 9.88 %	Yes – 90.9 % No – 9.1 %
6	Paperless Billing	Yes – 53.6 % No – 46.4 %	Yes – 74.9 % No – 25.1 %

Table 2. Description of identification signs that have three answer options.

No.	Identification sign	Churn customers	Non Churn customers
1	The indicator of the presence of several communication lines at the client (MultipleLines)	Yes – 41 % No – 49.12 % No phone service – 9.88 %	Yes – 45.5 % No – 45.4 % No phone service – 9.1 %
2	The type of communication line wire that the client is using (InternetService)	DSL – 37.9 % Fiber optic – 34.8 % No – 27.3 %	DSL – 24.6 % Fiber optic – 69.4 % No – 6 %
3	Customer's Internet Security Score (OnlineSecurity)	Yes – 33.3 % No – 39.4 % No internet service – 27.3 %	Yes – 15.8 % No – 78.2 % No internet service – 6 %
4	An indicator of whether the customer is backing up (OnlineBackup)	Yes – 36.8 % No – 35.9 % No internet service – 27.3 %	Yes – 28 % No – 66 % No internet service – 6 %
5	Indicator, the presence of device protection (DeviceProtection)	Yes – 36.3 % No – 36.4 % No internet service – 27.3 %	Yes – 29.2 % No – 64.8 % No internet service – 6 %
6	Indicator whether the user has technical protection (TechSupport)	Yes – 33.5 % No – 39.2 % No internet service –27.3 %	Yes – 16.6 % No – 77.4 % No internet service –6 %
7	The indicator of whether the client has live television broadcasts (StreamingTV)	Yes – 36.6 % No – 36.2 % No internet service – 27.3 %	Yes – 43.6 % No – 50.4 % No internet service – 6 %
8	The indicator of whether the client has streaming movies (StreamingMovies)	Yes – 37.1 % No – 35.6 % No internet service – 27.3 %	Yes – 43.8 % No – 50.2 % No internet service – 6 %
9	Type of the concluded payment agreement (Contract)	Month-to-month – 43 % One year – 25.3 % Two year – 31.7 %	Month-to-month – 88.6 % One year – 8.88 % Two year – 2.52 %

Table 3. Description of identification signs that have four answer options.

No.	Identification sign	Churn customers	Non Churn customers
1	Client payment type (PaymentMethod)	Mailed check - 25.1 % Electronic check – 25.1 % Credit card (automatic) – 25 % Bank transfer (automatic) - 24.8%	Mailed check – 16.5 % Electronic check – 57.3 % Credit card (automatic) – 12.4 % Bank transfer (automatic) – 13.8%

Also, data on the period of use of the services of a telecommunications company were used as input information, the parameter is calculated in months (tenure); the size of the client's monthly fee (MonthlyCharges) and the final amount of payments for the entire period of work with the client (TotalCharges). For each set of input information, consisting of 20 indicators, there is a predetermined output - Churn - characterizing whether the client left the telecommunications company with these parameters or not.

3 Modeling machine learning methods

The following machine learning methods have been selected for training, which have proven themselves well in solving various classification problems: AdaBoost - adaptive boosting; Decision Tree - decisive trees; Extra Tree Classifier - random trees; Gradient Boosting - gradient boosting; KNeighbors - k-nearest neighbors method; Logistic Regression - logistic regression; Naive Bayes - naive Bayesian classifier; Neural Network - neural networks; Random Forest - random forest method; SVM - support vector machine; XGB - Gradient boosting on trees.

To assess the effectiveness of the models, the Area Under the Curve (AUC) characteristic was chosen, a statistical indicator that is often used in machine learning methods that determines the area bounded by a certain curve and an abscissa axis, called the ROC curve (from receiver operating characteristic). The use of AUC for binary classification problems is popular because of its simplicity, intuitive interpretation [9], and also its use in the case of unbalanced datasets [13]. For a random classifier, the AUC is 0.5, and for an ideal classifier, the AUC is 1 [14]. Since there are only two options for a conclusion in the problem being solved (whether the client leaves or not), it is the ROC analysis, which is a graphical method for assessing the quality of the work of a binary classifier, that is most promising for assessing the proposed method for predicting the churn of customers of a telecommunications company. To construct the ROC curve, a pair of the following values is used: sensitivity and specificity. The value "sensitivity" characterizes the share of true-positive classifications in the total number of positive observations and is marked along the vertical axis of the ROC-curve graph, and the value "specificity" characterizes the proportion of true-negative classifications in the total number of negative observations and is marked along the horizontal axis of the ROC-curve graph. It should be noted that the higher the “sensitivity” value, the more reliably the classifier recognizes positive examples, and the higher the “specificity” value, the more reliable the classifier recognizes negative observations. Thus, the ROC curve reflects the relationship between the probability of false alarms (proportion of false-positive classifications) and the probability of “correct detection” (proportion of true-positive classifications). With an increase in sensitivity, the reliability of recognition of positive observations increases (the probability of "missing a target" decreases), but at the same time the probability of a false alarm increases. Table 4 shows the learning outcomes of individual machine learning methods and their characteristics according to the AUC metric.

Table 4. Qualitative assessment of predicting customer churn using machine learning methods.

No.	Machine learning method	AUC value
1	LogisticRegression	0.8038
2	AdaBoostClassifier	0.8024
3	XGB	0.8024
4	GradientBoostingClassifier	0.7962
5	RandomForest	0.7815
6	ExtraTreesClassifier	0.7725
7	SVM	0.7687
8	KNeighbours	0.7659
9	Neural Network	0.763
10	Naive Bayes	0.7602
11	DecisionTree	0.7408

Table 4 shows that three machine learning methods give the best results: LogisticRegression, AdaBoostClassifier, and XGB. For example, logistic regression gives correct results in 80.38% of examples. The authors decided to carry out computer modeling to identify ensembles of methods that optimally solve the problem of predicting the outflow of customers of a telecommunications company. For this, various combinations of the above machine learning methods were considered (Table 5 shows a fragment of the results obtained).

Optimization of method ensembles according to the criterion of the maximum AUC value allowed us to identify the best ensemble of methods containing such machine learning methods as Logistic Regression, Gradient Boosting and XGB, while the number of correct conclusions such an ensemble produces in 81.37% of examples.

4 Conclusion

Thus, as a result of the work carried out, the authors investigated 11 different machine learning methods to solve the problem of predicting the churn of customers of a telecommunications company. The best machine learning method was LogisticRegression, which showed an AUC of 80.38%. The use of Logistic Regression in conjunction with other machine learning methods within the ensemble of machine learning methods allowed us to increase the forecasting quality to 81.37. Further research by the authors will be devoted to identifying those identification parameters that significantly affect the process of predicting customer churn in order to use a smaller number of identification features with constant values of the forecast quality.

Table 5. Results of computer modeling of ensembles of machine learning methods.

No.	1 ensemble machine learning method	2 ensemble machine learning method	3 ensemble machine learning method	AUC value (%)
1	Logistic Regression	Gradient Boosting	gradient boosting on trees (XGB)	81.37
2	gradient boosting on trees (XGB)	adaptive boosting (AdaBoost)	Gradient Boosting	80.71
3	k-nearest neighbors method (KNeighbors)	gradient boosting on trees (XGB)	neural networks	80.62
4	Logistic Regression	Gradient Boosting	k-nearest neighbors method (KNeighbors)	80.23
5	Logistic Regression	Gradient Boosting	support vector machine (SVM)	80.2
6	Logistic Regression	Gradient Boosting	random forest	80
7	random forest	support vector machine (SVM)	neural networks	79.9
8	k-nearest neighbors method (KNeighbors)	Logistic Regression	naive bayes classifier	79.43
9	k-nearest neighbors method (KNeighbors)	naive bayes classifier	neural networks	79.43
10	k-nearest neighbors method (KNeighbors)	naive bayes classifier	gradient boosting on trees (XGB)	79.34

References

1. Verbeke, W. et al: Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert systems with applications, 38(3), 2354-2364 (2011).
2. Qureshi, S. A. et al: Telecommunication subscribers' churn prediction model using machine learning. In Eighth International Conference on Digital Information Management (ICDIM 2013), 131-136. IEEE, Islamabad (2013).
3. Balle, B. et al.: The architecture of a churn prediction system based on stream mining. Artificial Intelligence Research and Development, 256, 157-166 (2013).
4. Jain, H., Khunteta, A., Srivastava, S.: Churn Prediction in Telecommunication using Logistic Regression and Logit Boost. Procedia Computer Science 167, 101-112 (2020).
5. Bahrami, M, Bozkaya, B., Balcisoy, S.: Using Behavioral Analytics to Predict Customer Invoice Payment. Big Data, 8(1), 25-37 (2020).

6. Xie, S.-M.: Comparative models in customer base analysis: parametric model and observation-driven model. Journal of Business Economics and Management, 21(6), 1731-1751 (2020).
7. Kostić, S. M., Simić, M. I., Kostić, M. V.: Social Network Analysis and Churn Prediction in Telecommunications Using Graph Theory. Entropy, 22(7), 753 (2020).
8. Wassouf, W.N. et al: Predictive analytics using big data for increased customer loyalty: Syriatel Telecom Company case study. Journal of Big Data, 7, 29 (2020).
9. Baesens, B. et al: Profit Driven Decision Trees for Churn Prediction. European journal of operational research, 284(3), 920–933 (2020).
10. Faris H.: A Hybrid Swarm Intelligent Neural Network Model for Customer Churn Prediction and Identifying the Influencing Factors. Information, 9(11), 288 (2018).
11. Alexandru C. et al: Propensity to Churn in Banking: What Makes Customers Close the Relationship with a Bank? Economic Computation & Economic Cybernetics Studies & Research, 54(2), 77-94 (2020).
12. Schaeffer S. E., Sánchez S. V. R.: Forecasting client retention — a machine-learning approach. Journal of Retailing and Consumer Services, 52, 101918 (2020).
13. Ahmad, A.K., Jafar, A., Aljoumaa, K.: Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data, 6(1), 28 (2019).
14. Zhu, B. et al: Benchmarking sampling techniques for imbalance learning in churn prediction. Journal of the Operational Research Society, 69(1), 49-65 (2018).