Bulletin of the *Transilvania* University of Braşov • Vol 13(62), No. 1 - 2020
Series III: Mathematics, Informatics, Physics, 347-362
https://doi.org/10.31926/but.mif.2020.13.62.1.26

CUSTOMER CHURN PREDICTIVE MODELING BY CLASSIFICATION METHODS

Oleksandr DOROKHOV*, [2], Liudmyla DOROKHOVA [3], Lyudmyla MALYARETS [4] and Iryna USHAKOVA [1]

Abstract

The article describes methods of construction of predictive models for classifying customers based on their churn from the company for the example of a mobile operator. There are roles and tasks of customer analytics for understanding the business behavior of customers. The specificity of customer churn for companies associated with a subscription and transactional business model, involving regular customer payments is discussed, and the main reasons for churn are shown. Particular attention is paid to the analysis of forecasting methods based on classification methods. Here we discuss the forecast models based on the decision tree method and the Bayesian network. The decision tree method is basing on the C5.0 algorithm. The Bayesian model is constructed for a Naive and Markov structure. Customer service has become a key factor in the customer churn in all three models. A comparative analysis of the models was conducted based on indicators AUC and Gini. The decision tree model showed the best results. Moreover, the decision tree model shows the reasons why the customer can leave the company and give information for an individual approach to each customer. SPSS Modeler was used as a tool for building models.

2000 *Mathematics Subject Classification:* 90B50, 91B50, 62C12, 62C10
Key words: customer churn, classification methods, decision tree, bayesian nework.

1 Introduction

The success of any business is determined by the strategy of working with customers. The right strategy can lead to an increase in the number of customers.

[2] *Corresponding author*, Simon Kuznets Kharkiv National University of Economics, Ukraine, e-mail: aleks.dorokhov@meta.ua
[3] National University of Pharmacy, Ukraine
[4] Simon Kuznets Kharkiv National University of Economics, Ukraine
[1] Simon Kuznets Kharkiv National University of Economics, Ukraine

Satisfaction of the needs of each customer is the key to ensuring the competitiveness of the company.

It becomes more and more urgent to know your customer, to have a complete idea of what and when they bought, how satisfied they are with the cooperation. This information allows us to predict buying behavior and determine the prospects for further development of relations.

Currently, the tasks of analyzing customer behavior are central to the analytical reports of many companies. Increasingly, terms such as customer experience, customer understanding using in reports.

According to Gartner's report "CMO Spend Survey 2017-2018: Budgets Recede Amid Demand for Results" marketing budgets in 2017 remained at the level of the previous year. At the same time, the analyst was allocated the largest pro-cent of budgets (9.2%).

In the report "2017 Global Customer Experience Benchmarking Key Findings Report" from Dimension Data, 82% of companies consider customer experience as one of the main differences from competitors, and 77.5% define it as the main indicator of strategy effectiveness.

The "Data Elevates the Customer Experience" report from Forbes Insights presents the results of a survey of analysts from different organizations, where 42% of respondents indicated that the data would significantly affect the quality of customer experience in the next 2 years [8].

To form a strategy of relations with customers, the company collects and analyzes data on customer behavior [7]. Customer analytics is a rapidly developing direction, the essence of which is the collection and processing of customer information for segmentation and forecasting the features of their behavior, making informed business decisions [6].

Analytics is an expensive intellectual resource for achieving business goals, so it should be used where it clearly brings money (due to the flow of customers, increasing market share, etc.) or saves resources. Technologies have been developed in this field and have become relevant in Ukraine. In addition, these technologies have established themselves in the West.

Customer analytics describes the customer, shows his portrait, needs and predicts his behavior. Analytics associated with instrumental data mining and other special tools for data analysis [9]. Most of the tasks of such an analysis are reduced to the prediction of certain events.

The most popular are the forecasts for the purchase of certain products, the response to a certain marketing campaign or the customer churn from the enterprise, that is, the termination of the use of its services.

The technical result of such a forecast is an estimate of the probability of occurrence of an event. In modern conditions of high competition, selling a new service to an existing customer is several times cheaper than attracting a new one. The main task today is not to attract new customers, but to retain existing ones. Therefore, special attention is paid to building profitable and long-term relationships with existing customers, forecasting and finding out the reasons for the customer churn.

The purpose of this work is to build a predictive model of customer churn for companies that conduct their business on the basis of subscribing customers, using machine learning methods.

To achieve the goal it is necessary: to review and describe the trends of customer analytics, to analyze approaches and methods for forecasting customer churn, to build a predictive model of customer churn using machine learning methods, to analyze the adequacy and applicability of the model.

2 Approaches and methods of forecasting customer churn

Customer churn is the loss of customers, which results in a lack of purchases or payments over a period of time. The churn is related to customer's reluctance to renew the contract, make advance payments, and the closure of the bill or account.

Customer churn can significantly affect the company's revenues. The analysis of customer churn applies only to those areas of activity where purchases are of a regular nature. This indicator is especially important for companies with a subscription and transactional business model, involving regular payments of customers.

This area includes banks, telecom operators, Internet service providers, cloud services, and so on. The following types of churn are distinguished, they are shown in Figure 1.

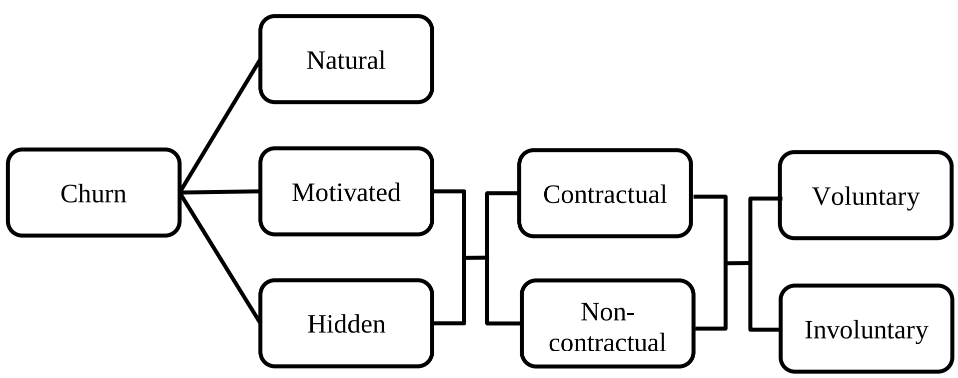

Figure 1. Customer churn

Natural churn does not depend on the activities of the company (change of place of residence, death, changing needs, etc.). Motivated churn - the customer refused the services of the company for the benefit of another company or found an alternative way of satisfying his needs.

Hidden churn - the customer continues to use the services of the company, but the volume and regularity of consumption is reduced. The last type of churn is a sure sign that the customer is simultaneously using the services of competitors or substitute products.

The latent churn is usually not considered a churn, although most of the lost customers first reduce the share and regularity of consumption before finally breaking the relationship with the company.

Depending on the form of customer service, the churn can be for a contractual or non-contractual basis. In the first case, the customer makes purchases at certain times in accordance with the contract or by automatic payment.

Customer care is clearly observed and fixed. For example there are mobile operators. In the second case, customer commit and pay for purchases himself and at any time. The churn of the customer is not observed explicitly. The main example is shopping in retail.

Depending on the procedure for completing the interaction, care can be voluntary or forced. In the first case, the customer simply stops using the service. In the second case, the customer is forced to stop servicing or payment. We will give an example of the expiration of the credit card.

Churn is motivated and observed at the termination of the contractual relationship, which is often based on the so-called "subscription". Interactions are not always related to the contract, especially in need of timely forecasting customer churn. In this case, the task is to determine the time point after which the return of the customer will be impossible or too expensive.

This point is usually in the period of decreasing activity or inaction of the customer and is used to predict the churn event [2, 11]. The task of the customer analyst in this case is reduced to predicting the possibility of the customer getting into the churn group, further segmenting the customers in terms of their profitability level and working with those customers that bring the greatest revenue.

The reasons for customer churn may differ depending on the scope of activity and the specifics of the company. However, we can identify a number of common reasons for customer churn:

dissatisfaction with the quality of the company's products and services;

low level of customer service;

more attractive offers of competitors;

the emergence of products or services of substitutes;

decrease in brand loyalty.

Various methods are used to predict customer churn: statistical methods, classification methods based on machine learning, simulation methods [1, 3, 4, 10 - 13].

Using these methods, the historical data on purchases and behavior are processed to predict the chances of churn for each customer. Recently, the classification methods based on machine learning are very popular.

In most of the studies examined, several models are compared on a single data set. This is a good practice for building a classifier.

An overview of some methods of data mining: neural networks, statistical methods, decision trees and coverage algorithms and their use in the context of customer churn analysis is given to the paper [1].

The article [2] explores the predictive modeling of customer churn for a company specializing in the delivery of products by subscription.

For modeling, methods such as logistic regression, decision trees, neural networks, and SAS VDMML tools are used.

It is concluded that the choice of method depends on the goal - the accuracy of the model or the correct classification.

In paper [4] the main emphasis is on using the Bagging method to create several versions of the classifier and to obtain an aggregated classifier on their basis.

The study focuses on the support vector method, using the example of distribution, subscribes to a subscription [5].

A comparison is made between the two methods of choice of parameters necessary for the implementation of the method of free vectors.

This method is compared with logistic regression and random forests.

It is said that only when applying the procedure for selecting the optimal parameter, the support vector method exceeds the traditional logistic regression, while the random forests exceed the support vector method for both parameters.

Also in detail research [4] the churn is considered to depend on the customer's profile and the financial costs of its retention are taken into account.

The are compared such classification algorithms: Bayesian algorithm, logistic regression, decision trees. Finally, the best results showed a Bayesian approach.

The methods considered in the work give different results depending on the context of use.

They depend on the method of data collection, their structure, preliminary processing, and selection of the best parameters for the models.

However, we can identify a number of methods that give the best results for predictive modeling of the customer churn: logistic regression, decision trees, random forest, neural networks and Bayesian algorithm.

The simulation results depend not only on the chosen method, but also on the algorithm used in it [3].

A comparison of the properties of the algorithms for the two-level classification is presented in Table 1.

Currently, new improved classification algorithms have been developed, as for example C5.0 [3].

Table 1.

Two-class classification algorithm properties

Algorithm	Accuracy	Training time	Linearity	Parameters
Logistic regression		fast	yes	5
Decision forest	high	moderate		6
Decision jungle	high	moderate		6
Boosted decision tree	high	moderate		6
Neural network	high			9
Averaged perceptron	good	moderate	yes	4
Support vector machine		moderate	yes	5
Locally deep support vector machine	good			8
Bayes' point machine		moderate	yes	3

The actual use for solving the above task is fraught with considerable financial costs and can greatly affect the profitability of the company both in the positive and in the negative direction.

Therefore, there arises the problem of choosing the most effective algorithm, the prognostic capabilities of which are the main criterion of choice. To solve this problem in the field of mobile communication services, various types of algorithms for data mining technology can be used.

3 Used tools, methods, algorithms for classification

There are large amounts of data and complex algorithms that are used, and can be realized through the use of information technologies and tools in customer analytics. One of the leaders in the field of customer analytics is IBM SPSS Modeler.

SPSS Modeler is a predictive analytics platform that helps you quickly develop accurate forecasting models and apply predictive analytics at the level of individual users, groups, systems and the whole enterprise. It includes a number of improved algorithms and methods.

Their actual use to predict customer churn can greatly affect the profitability of the company. Therefore, there arises the problem of choosing the most effective algorithm, the prognostic capabilities of which are the main criterion of choice.

For predictive modeling of customer churn in work, two methods will be used: the decision tree and the Bayesian network, which will be implemented by tools in SPSS Modeler.

3.1 Decision tree

Decision tree models allow you to create classifications that you can use to predict events based on a set of decision rules. Separation of existing data into classes based on specified characteristics can be used to construct rules.

These rules can be used to classify observations with maximum accuracy. For example, you can build a tree that classifies customers based on their propensity to churn.

This approach has several advantages. The process of reasoning behind the model is clear. Its internal logic can be understood by looking at the tree.

The process automatically includes in its rules only those attributes that actually matter when making a decision. Attributes that do not contribute to the accuracy of the tree are ignored.

The decision tree model can be represented in two forms: as a set of rules IF-THEN or as a tree. The representation in the form of a set of rules is more compact. It allows us to understand how specific groups of elements relate to a concrete conclusion.

The tree view more clearly shows how the attributes in the data can help break up or split the entire population into subsets that are relevant to the problem.

3.2 Automatic classification algorithms

For the classification by the decision tree method, you can use different algorithms. The most popular are C4.5 and its newer version of C5.0. Algorithm C5.0 is faster and gives better classification results compared to C4.5. Currently, the C5.0 algorithm is in fact the standard for building decision trees.

This algorithm has the following advantages:

performs a classification several orders of magnitude faster than C4.5;

uses less memory than C4.5;

gets results similar to C4.5, but with significantly smaller decision trees;

it is universal, it solves well the classification for different subject areas;

it is used to analyze not only numerical, but also nominal data;

it provides processing of the missed data;

it selects from the set of attributes only those that affect the classification result;

it supports boosting for better and more accurate trees;

it allows you to weigh various cases and types of incorrect classification.

3.3 Bayesian Network

Bayesian network is a graphical model in which the variables $x = \{x_1, x_2, ..., x_p\}$ from the set of random variables are represented by the nodes of the graph $G = (V, E)$, each node $v_i \in V$ corresponds to a random variable x_i, which are probabilistic, or conditionally independent of each other.

Links, or arcs, between the nodes of the Bayesian network sometimes, but not always, correspond to cause-effect relationships. The training of the Bayesian network consists of two stages: structural training and parametric training.

Structural training of the Bayesian network consists of determining the structure of the graph while the parametric training - consists of determining the parameters of each node. This information can be obtained directly from the data.

Networks have a very high resistance to missing information and give the best possible predictions based on available information.

Advantages of using Bayesian network:

allows to learn about the cause-effect relationships; thanks to this, it becomes possible to study the problem area and predict the consequences of the changes that have appeared;

it provides an effective approach to prevent the retraining of data;

it provides a convenient, visual representation of the relationship.

4 Predictive modeling of customer churn of the mobile operator

Mobile operators have a business model based on subscription. Forecasting customer churn is especially important for such a business. Next, a predictive modeling of the customer churn for the company providing mobile communication services will be performed.

4.1 Data preparation

The construction of the predictive model in SPSS Modeler has features related to the graphical interface. It is based on the representation of nodes and threads.

Nodes are symbols representing individual operations on data.

The nodes are connected by links. Links indicate the direction of data movement in the stream. A thread is a process in the form of a sequence of operations on data.

Typically, a thread is used to retrieve data from a set of data sources, perform operations on that data, and send the results to an output object, which can be a table or view.

When constructing a predictive model of the customer churn on the basis of the learning classification, it is necessary to create a flow for preliminary processing of data related to obtaining, training, and preparing data for classification.

This stream includes nodes: data source, data separation, output table and data types. The data source is a database of customers (5 thousand records), the ratio of the remaining customers is 85% to 15%.

Customer information contains the following attributes: region, user experience, area code, phone number, international plan, total number of minutes, total number of minutes, total eve , total night charges, total international calls and the total international calls, total international calls, number of customer service calls churned.

The Partition node divides the data into two subsets of 1_Training and 2_Testing in a ratio of 70% and 30%.

The Table node is designed to store the split results.

The Type node specifies the properties of the fields. Here the role of the fields is defined: the field is input or target for modeling.

The target field is churn. At this stage, the phone number field was disabled.

4.2 Construction of customer churn forecast models

The construction of a predictive model of customer churn includes: selection of a model node, description of model parameters, creation of a model snapshot.

A model nugget is a set of rules, formulas, or equations that represent the results of model building operations.

The purpose of the snapshot is to scrape the data to create a prediction and further analyze the properties of the model.

To evaluate the capabilities of the obtained models, two Analysis and Evaluation nodes will be used to generate accurate predictions.

The analysis performs various comparison operations between the forecast value and the actual values of the variables for one or more model casts.

In addition, using Analysis, comparisons are made between some forecast models and others. Evaluation allows you to visually compare diagrams with the help of diagrams, and then choose the best model.

In our case, the predictive modeling of the customer churn is carried out on the basis of the churn signs: remained (False), left (True) for two subsets of the Learning and Test data. For modeling, methods such as decision trees and Bayesian networks will be used.

The modeling stream for the Decision Tree model is shown in Figure 2. To build the model, algorithm C5.0 was chosen.

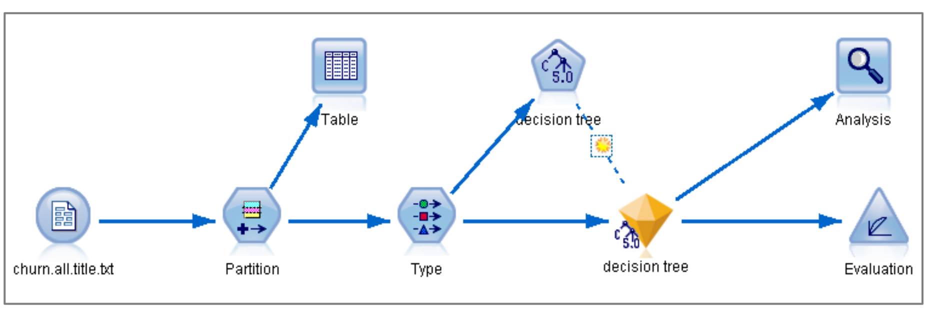

Figure 2. The modeling stream for the decision tree method

Prior to the construction of the model, the validity of the predictors is checked. Predicators, which turn out to be irrelevant for verification, are excluded from the process of constructing the model.

After completing such a check, the fields were included in the model: number of customer service calls, total evening charge, total evening minutes, total number of minutes, total number of minutes.

To increase the accuracy of the model, a busting was used. Busting works by consistently creating several models. The first model is constructed in the usual way.

Then, when building the second model, those records that were incorrectly classified by the first model and so on are processed. Eventually, observations are classified using the entire set of models and using a weighted voting procedure to combine individual predictions into one general prediction.

When constructing the model, a truncation of the tree was used. In the first stage, subtrees were checked, and in the second stage the tree was considered as a whole, and the weak branches were folded.

As a result of the construction of the model, nine rules were obtained. The main factor in the customer churn was the dissatisfaction with the quality of service (number customers service call), the total duration of calls (total day minutes). The constructed model of the decision tree is shown in Figure 3.

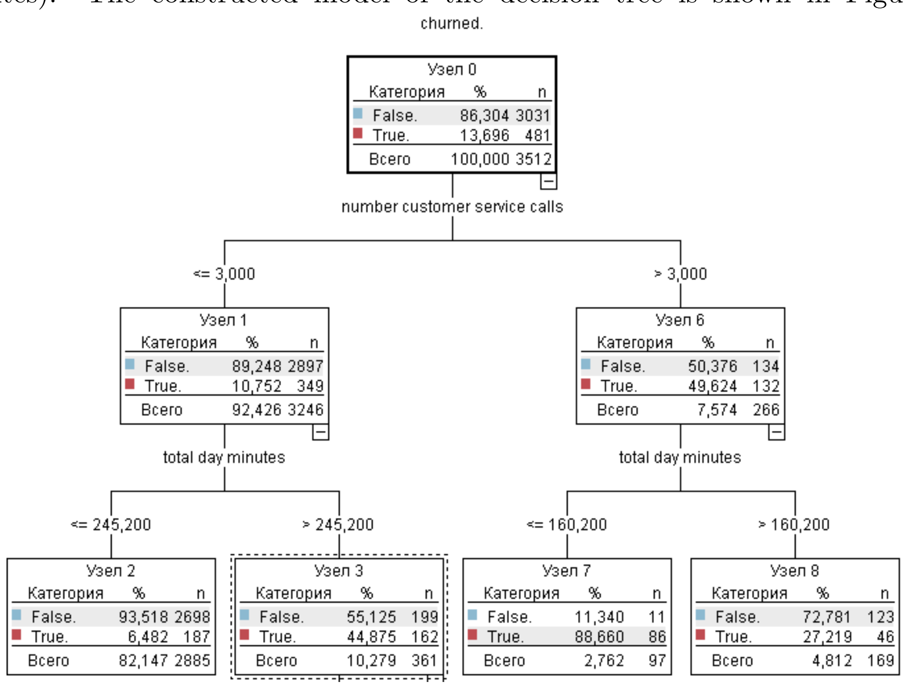

Figure 3. Decision tree

Using this model, you can predict the customer's entry into the churn group and find out the reasons for the churn. In the future, you can use this information to keep the customer.

To obtain a more accurate model, two structures of the Bayesian network were selected, as shown in Figure 4).

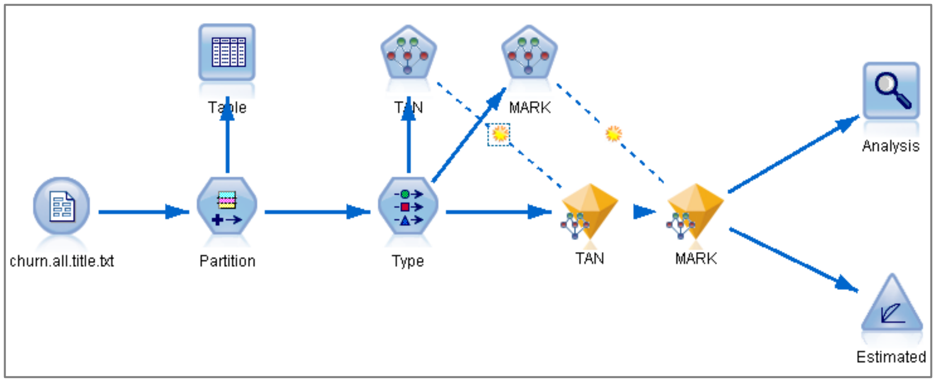

Figure 4. Modeling stream for Bayes network

TAN is a lively Bayesian tree model (Tree Augmented Naive Bayes model), which creates a model of a simple Bayesian network, which is an improvement to the standard naive Bayesian model.

This is achieved by the fact that each predictor can depend on another predictor, and not only on the variable assignment, which increases the accuracy of the classification;

MARC is a markup structure that selects the following set of nodes in the data set: the parents of the destination variable, the child nodes of the destination variable, and the parents of the child nodes.

This method of building a network is considered more accurate, but for large data sets it is more time consuming.

After adjusting the model parameters, the model snaps were created. Naive Bayesian network has constructed a classification with all variables taken into account, as shown in Figure 5.

Three variables were chosen when using Markov structure: international plan (wait = 0.5), number customers service call (wait = 0.44), voice mail plan (wait = 0.06), as shown in Figure 6.

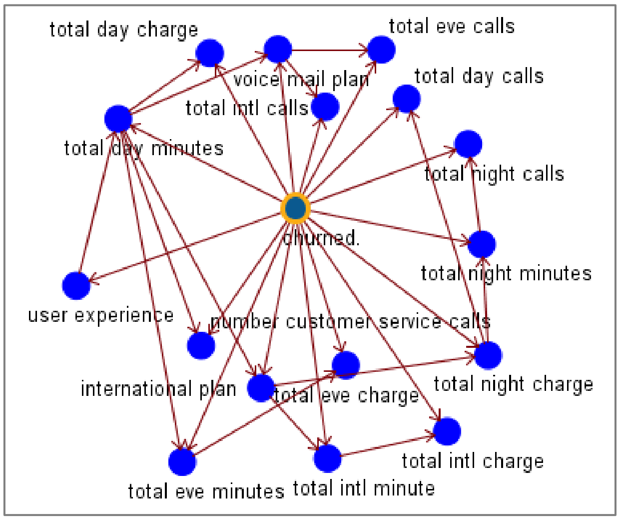

Figure 5. Naive Bayesian network to predict customer churn

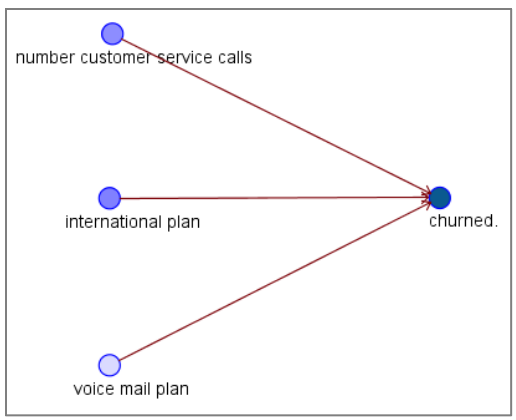

Figure 6. Markov structure to predict customer churn

5 Analysis of results

Originally a choice was made for the Bayesian model. The reports for comparing the accuracy of the obtained models are presented in Tables 2-3.

In the first case, the number of correct forecasts is 91.46% for 1_Training and 88.98% for 2_Tasting. In the second, it is 85.62% and 86.85% respectively. Therefore, naive Bayesian network was chosen.

Table 2.

Comparison of \$ B-churned with churned

Subsets	1_Training		2_Tasting	
Correct	3 212	91,46%	1 324	88,98%
Wrong	300	8,54%	164	11,02%
Tonal	3 512		1 488	

Table 3.

Comparison of \$ B1-churned with churned

Subsets	1_Training		2_Tasting	
Correct	3 050	86,85%	1 274	85,62%
Wrong	462	13,15%	214	14,38%
Tonal	3 512		1 488	

Next, we perform a comparison of the simulation results by the Decision Tree method and the Bayesian network. To compare the characteristics of the obtained models, an ROC (receiver operating characteristic) diagram was constructed, as shown in Figure 7.

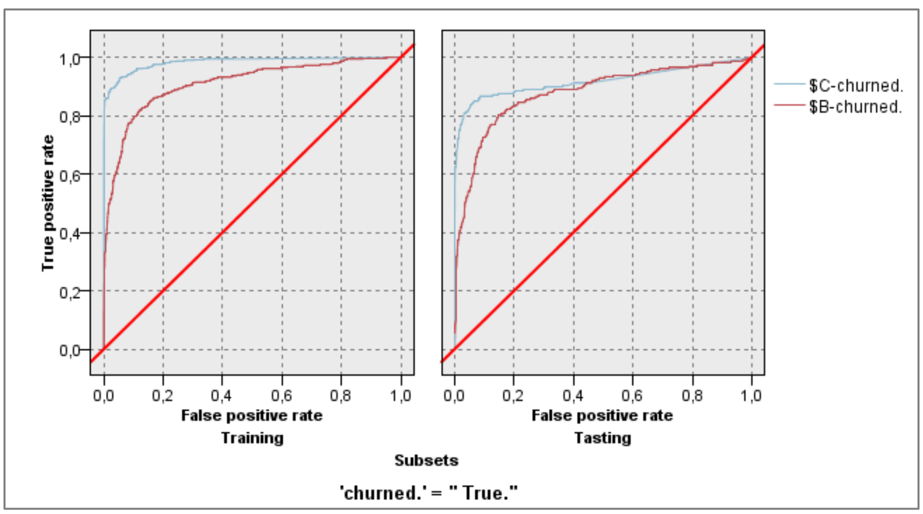

Figure 7. ROC diagram for decision tree model and Bayesian model

This diagram is used to visualize, organize and select classifiers based on their effectiveness. The ROC diagram shows the dependence of the share of true positive conclusions on the share of false positive conclusions of the classifier.

The ROC diagram displays the ratio between the fraction of objects correctly classified and the proportion of objects that are erroneously classified as bearing a characteristic when the threshold of the decision rule is varied.

When visualizing the ROC curves, their location relative to each other indicates their comparative effectiveness. The curve located above and to the left, testifies to greater predictive ability of model.

Thus, in Figure 7 two ROC-curves are combined on one graph. It is seen that on both subsets of the data, the decision tree model ($ C-churned) as a whole shows better results than the Bayesian network ($ B-churned).

Visual comparison of the ROC curves does not always make it possible to identify the most efficient model. The quantitative interpretation of ROC is given by AUC, the area bounded by the ROC curve and the axis of the fraction of false positive classifications.

The higher the AUC, the better the classifier, while the value 0.5 demonstrates the unfitness of the chosen classification method (corresponds to a random choice).

A value of less than 0.5 says that the classifier operates exactly the opposite: if the positive ones are called negative and vice versa, the classifier will work better.

The Gini coefficient is calculated as the double area between the ROC curve and the diagonal (or as Gini = 2AUC-1). The Gini coefficient is always between 0 and 1, and the higher it is, the better the classifier.

In the unlikely event that the ROC curve is below the diagonal, the Gini coefficient will be negative. Table 4 shows the Evaluation Metrics AUC, Gini for the decision tree and the Bayesian model.

Table 4.

Evaluation metrics (AUC, Gini)

Subsets	1_Training		2_Tasting	
Model	AUC	Gini	AUC	Gini
$C-churned	0,983	0,967	0,918	0,835
$B-churned	0,91	0,82	0,878	0,756

The model of the decision tree has better results. Therefore, in order to forecast the customer churn from a mobile operator, it is better to use the decision tree model.

6 Conclusions

Currently, business conditions are associated with fierce competition. Therefore, many companies use customer analytics to enhance their competitive advantages.

One of the modern business models is the subscription of customers. For such a model, an urgent task is to reduce customer churn. Timely retention of the customer requires several times less costs than attracting a new one.

At the disposal of the business, the customer analyst provides tools to analyze and predict the churn for the timely retention of customers. A well-constructed model can provide information for a wide range of solutions.

To date, the most common methods are based on machine learning. These methods are based on the use of complex classification algorithms, so the tools of information technology are used to implement them in customer analytics.

In this paper, we discussed the construction of predictive models of customer churn for the mobile operator based on the decision tree methods and the Bayesian network. The Bayesian model was built in two versions: a naive and a Markov network.

A more accurate forecast was shown by the naive Bayesian network, so in future the predictive models of the decision tree and Bayesian naive network were compared. In general, the best result was obtained for the predictive model of the decision tree.

The analysis of the received models has shown that in all models the main factor of the customer churn is the number of calls to the support service. This indicates that the constructed models reflect the essence of the processes occurring in business, and can be recommended for use.

Indeed, frequent customer contact with the customer service can be the main reason for the customer's loyalty reduction. And at this rate, companies need to pay attention in the first place.

The decision tree model provides additional benefits. The rules built on its basis show the reasons why the customer can leave the company and give information for an individual approach to each customer.

References

[1] Almana, A., Aksoy, M., Alzahrani, R.,*A survey on data mining techniques in customer churn analysis for telecom industry*, International Journal of Engineering Research and Applications **4** (2014), 165-171.

[2] Bahnsen, A., Aouada D., Ottersten B.,*A novel cost-sensitive framework for customer churn predictive modeling*, A Springer Open Journal Decision Analytics **2** (2015), no. 1, 1-15.

[3] Berry, M., Linoff, G.,*Data mining techniques: for marketing, sales, and customer relationship management*, Wiley (2011).

[4] Canale, A., Lunardon, N.,*Churn prediction in telecommunications industry. A study based on bagging classifiers telecom*, Collegio Carlo Alberto Working Papers (2014).

[5] Coussement, K., Poel, D.,*Churn prediction in subscription services: an application of support vector machines while comparing two parameter-selection techniques*, Expert Systems with Applications **34(1)** (2008), 313-327.

[6] Dorokhov, O., Dorokhova, L., Zorina, E., *A fuzzy approach and modelling of service estimations for drugs freight transportation*, Transport and Telecommunication **11(1)** (2010), 19-25.

[7] Dorokhova, L., Dorokhov, O., *Computer fuzzy model regarding pharmacies integral perceptions by visitors*, Bulletin of the Transilvania University of Brasov, Series III: Mathematics, Informatics, Physics **10(2)** (2017), 155-170.

[8] McIntyre, E., Virzi, A., *CMO spend survey 2017-2018: budgets recede amid demand for results*, Gartner (2017).

[9] Omelchenko, O., Dorokhov, O., Kolodiziev, O., Dorokhova, L., *Fuzzy modeling of the creditworthiness assessments of banks potential borrowers in Ukraine*, Ikonomicheski Izsledvania **27(4)** (2018), 100-125.

[10] Ozerov, A., Solovev, R., *Modeling taxi service customer behavior*, Cloud of Science **1(4)** (2014), 579-594.

[11] Pradeep, B, Rao, S., Puranik, S., Hegde, A.,*Analysis of customer churn prediction in logistic industry using machine learning*, International Journal of Scientific and Research Publications **7(9)** (2017), 401-413.

[12] Ushakova, I.,*Simulation of the impact social media on promoting education services*, CEUR Workshop Proceedings **1614** (2016), 575-590.

[13] Ushakova, I., Shevchenko, A.,*Methods for forecasting information products and services market*, Development Management **14** (2008), 142-145.