International Journal of
Intelligent Systems and Applications in Engineering

ISSN:2147-6799 www.ijisae.org Original Research Paper

Customer Satisfaction Using Data Mining Approach

Burcu ORALHAN *[1], Kumru UYAR[2], Zeki ORALHAN [3]

Accepted 3rd September 2016

Abstract: Customers and products are the main assets for every business. Companies make their best to satisfy customers because of coming back to their companies. After sales service related to different steps that make customers are satisfied with the company service and products. After sales service covers different many activities to investigate whether the customer is satisfied with the service, products or not? Hence, after sales service is acting very crucial role for customer satisfaction, retention and loyalty. If the after sales service customer and services data is saved by companies, this data is the key for growing companies. Companies can add value their brand value with the managing of this data. In this study, we aim to investigate effect of 6 factors on customer churn prediction via data mining methods. After sale service software database is the source of our data. Our data source variables are Customer Type, Usage Type, Churn Reason, Subscriber Period and Tariff The data is examined by data mining program. Data are compared 8 classification algorithm and clustered by simple K means method. We will determine the most effective variables on customer churn prediction. As a result of this research we can extract knowledge from international firms marketing data.

Keywords: *Data Mining, Customer Churn Prediction, Customer Satisfaction, Knowledge Discovery in Database*

1. Introduction

Telecommunication companies maintain and store tremendous amount of data about the customer information, phone calls and the operations of their networks. Due to the improvement of computer systems and telecommunication technologies, this industry has expanded rapidly. Data mining helps to improve quality of service, detect the customer communication type, determine deceitfulness activities and make better use of resource in telecommunication industry.

One of the widely applied areas of data mining is Customer Relationship Management. Customer Relationship Management approach is focuses on retention, relationship development and increase satisfaction. From a business intelligence perspective, churn management process under the customer relationship management (CRM) framework consists of two major analytical modelling efforts: predicting those who are about to churn and assessing the most effective way that an operator can react (including 'do nothing') in terms of retention [1]. Previous studies have indicated that the cost of gaining new customers is much expensive than the cost of retention the existing customers [1], [2], [3], [4], [5], [6] Loss of customer causes negative effects on firms reputations and income reduction [7]. Companies are increasingly focused on accurate customer churn forecasting models. These models should be enhanced to identify the factors to churn and the developed needs to retain customers. Churn of customer is defined as the tendency of customers to interrupt doing business with a firm in the time process. Customer churn has become a significant problem today. In worldwide, one of the main challenges is determined to propensity of churn of customers. Geppert (2003) suggest another list of causes of churn: Price, Service quality, Fraud, Lack of carrier responsiveness, Brand disloyalty, Privacy concerns, Lack of features, New technology introduced by competitor, New competitor enter the market, Billing or service disputes [8] .

High customer churn risk is one of the stringent challenges in telecommunication companies. Pareek has classified telecommunication industry challenges in four groups; Consolidation, Competition, Commoditization, Customer service. Telecommunication data which is very complex has needed many pre-processing for analyse. Data mining which is a useful tool, is extracting and exploring information from data. In this paper, we will provide an application to telecommunication companies. We want to have knowledge about customers that who are existing and loyal to company, who are going to leave or quitting from company products/services.

Customer churn probability is predicted by data mining algorithms analysing historical data. Data mining techniques used for this purposes typically utilize the average number of calls, billing data, the change in the average number of calls, call detail data, subscription and customer information [10].

In this paper, we investigate features of applying the data mining in customer churn prediction of a telecommunication data. First, we identify the variable of effecting customer churn prediction of a telecommunication system. These variables are examined in data mining algorithms. Data mining algorithms are compared with each other. Proposed algorithm is also discussed and we conclude the paper with conclusions.

2. Material And Method

Telecommunication data consist lot of interesting issues for

[1] *Faculty of Economics and Administrative Sciences, Nuh Naci Yazgan Uni. – 38030, Turkey*
[2] *Faculty of Economics and Administrative Sciences, Nuh Naci Yazgan Uni. – 38030, Turkey*
[3] *Turk Telekomünikasyon A.Ş. – 38104, Kayseri/ Turkey*
** Corresponding Author: Email: boralhan@nny.edu.tr*
Note: This paper has been presented at the 3rd International Conference on Advanced Technology & Sciences (ICAT'16) held in Konya (Turkey), September 01-03, 2016.

analysing in data mining. Telecommunication databases include huge amount of data in worldwide. But raw data is not usually suitable for direct usage for data mining. For this reason researchers should examine data effectiveness and apply some of pre-processing operations to be usable. Our study include of data sampling, pre-processing, model construction, and model evaluation phases.
We have used open-source data mining tool Weka (version 3.8) for this study. We analyse our dataset performance of a comprehensive set of classification algorithms (classifiers).

2.1. Data Set

The dataset that we used obtained from a Turkey's one of the biggest telecommunication company. Real time data was obtain on May 2016. Untreated data consisted of 6 factors, 66 sub factors and 498,866 instances. 38,36% of data were churners.
Evaluate factors is following;

Table 1. Variable and Codes

Variable	*Code*	*Variable*	*Code*
Customer Type	Active	Subscriber Period	0-5
	Churn		6-10
Usage Type	H		11-15
	B		16-
Churn Reason	1	Tariff	1
	2		3
	5		4
	6		7
	9		8
	17		14
	25		Other
	Other		

According to (Acker, et al., 2013), data pre-processing can be up to 80% of the total analysis work and analyzing the data, once joined, cleaned and transformed consumes just about 20%. The raw world data is usually noisy, inconsistent and incomplete due to their typically large size and their likely origin from multiple and heterogeneous sources. Dimensions of data quality; Accuracy, Completeness, Consistency Timeliness, Believability, Value added, Interpretability Accessibility.
Data sampling randomly selects a set of customers with the required information, according the definition of churn in designated telecommunication company.
Evaluate factors and sub factors distribution as shown in Fig. 1.

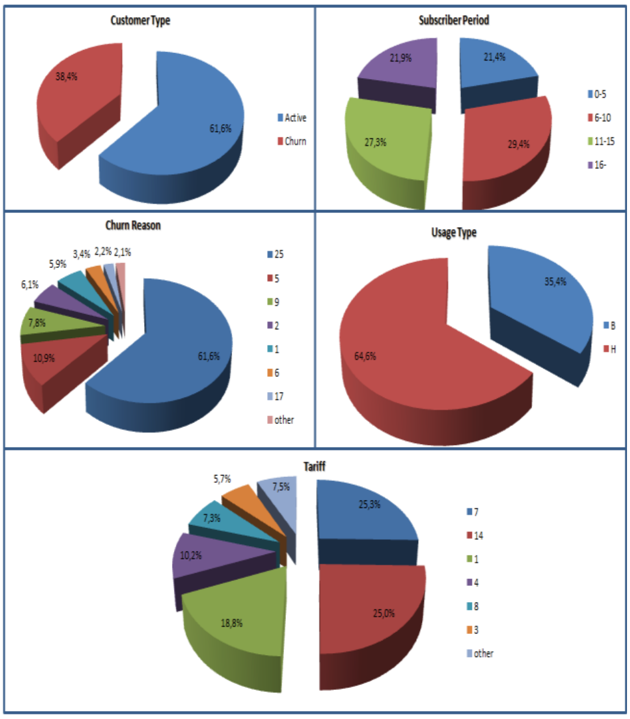

Figure 1. Distribution of the data

Data pre-processing eliminates the unrelated information which consists missing values, incorrect words, wrong mathematical symbols, duplicated information, and like that errors [2]. Major data pre-processing tasks and techniques are shown in Fig. 2.

Figure 2. Major data pre-processing tasks and techniques [12]

In the study we used some of pre-processing techniques. In data cleaning step we clean some of the data, which are identifying incorrect information or incorrect. At the end of this step, we have 498.356 data. Ordered sub factors are converted to numeric values. We create some of groups of subscribe period for subscriber time.

2.2. Data Mining Techniques in WEKA

Data mining technique in CRM is usually apply in real world case because CRM on data mining have attracted both the practitioners and academicians. In CRM there are several different functionalities, techniques and applications. as shown in Table 2.

Table 2. Data Mining Functionality, Technique and Application

Functionality	*Technique*	*Application*
Association	Set Theory	Cross Sell
	Statistics	
	Bayesian Classification	
Estimation	Neural network	Exchange rate estimation
	Statistics	
	Time series	
Classification	Decision Tree	Credit embezzle Market segmentation
	Fuzzy	
	Neural network	
	Genetic Algorithm	
Prediction	Regression	Churn Prediction Fraudster prediction
	Neural network	
	Decision Tree	
Segmentation	Neural network	Market segmentation
	Statistics	
	Genetic Algorithm	
	Decision Tree	

For churn prediction is used most popular ones are:

- Decision trees
- Artificial Neural networks
- Regression.

Weka (Waikato Environment for Knowledge Analysis) includes some of classifier algorithm such as Bayes, MISC, Functions, Rules, Decision Tree, Lazy etc. A good mix of algorithms have been chosen from these groups that include Naive Bayes and Bayes net from Bayes Classifier, Multilayer Perceptron from functions, JRip, PART, OneR from Rules Classifier and Random Forest and J48 from Trees.

3. Results and Discussion

A total of 8 classification algorithms have been analysed in this study. We evaluate the features with 10-fold cross validation. In the classification process, we perform 10 times for training and testing.

The results from TABLE III have been analyzed the classifiers work better. The classifiers J48, Random Forest and PART have performed better in data set (up to %78). Errors and Kappa statistic seem to be same among 3 classifiers and are based on the accuracy of the prediction.

Table 3. Comparison of Different Classifiers Using Telecom Data Set

Classifier	*Time taken to build model*	*Correctly classified instances*	*Kappa statistic*	*Mean absolute error*	*Root mean squared error*	*Relative absolute error*
J48	0,25	78.24 %	0.628	0.0662	0.182	44.68 %
Naive Bayes	0,06	77.92 %	0.6229	0.0653	0.1849	44.0 %
JRIP	9,72	65.22 %	0.1386	0.1362	0.261	91.91 %
PART	0,8	78.23 %	0.6279	0.0662	0.1819	44.64 %
Random Forest	14,75	78.23 %	0.628	0.066	0.1817	44.52 %
MultiLayer Perceptron	26,06	77.93 %	0.6232	0.0669	0.1826	45.15 %
BayesNet	2,07	77.92 %	0.6229	0.0653	0.1849	44.08 %
OneR	0,11	72.50 %	0.5247	0.0687	0.2622	46.36 %

Customer clustering is very important issues in data mining methodologies for customer relationship management (CRM).

To segment customers by Customer Type, Usage Type, Churn Reason, Subscriber Period and Tariff as variables and used K-Means to model the customers into five clusters. To generate roughly the same number of subscribers in each of the seven clusters. Table IV is summarized the clustering results. Simple K-Means method is a common and effective method for clustering.

Table 4. Cluster of Customer Segmentation

Cluster	*Full Data*	*0*	*1*	*2*	*3*	*4*	*5*	*6*
Attribute	498356	85008	109671	65748	86170	67292	62389	22078
%	100	17%	22%	13%	17%	14%	13%	4%
Active_ Churn	Active	Active	Active	Churn	Churn	Active	Active	Churn
H_B	H	H	H	B	H	H	H	B
Period	6-10	0-5	6-10	6-10	6-10	16-	11-15	0-5
Tariff	7	4	14	1	14	7	7	3
Churn_ Reason	25	25	25	1	5	25	25	2

From this result, it is possible to generate some customer management strategies. Simple algorithm K means the results can be interpreted as follows.

- Cluster 0: Include %17 of all customers and these customers are active in system 0-5 year period. They use 4.tariff in telecommunication systems for their home.
- Cluster 1: Include %22 of all customers and these customers are active in system 6-10 year period. They use 14.tariff in

telecommunication systems for their home.

•Cluster 2: Include %13 of all customers and these customers are churned in system 6-10 year period. They used 1.tariff in telecommunication systems for their business. Churn reason is 1 coded.

•Cluster 3: Include %17 of all customers and these customers are churned in system 6-10 year period. They use 14.tariff in telecommunication systems for their home. Churn reason is 5 coded.

•Cluster 4: Include %14 of all customers and these customers are active in system 16- year period. They use 7.tariff in telecommunication systems for their home.

•Cluster 5: Include %13 of all customers and these customers are active in system 11-15 year period. They use 7.tariff in telecommunication systems for their home.

•Cluster 6: Include %4 of all customers and these customers are churned in system 0-5 year period. They used 3.tariff in telecommunication systems for their business. Churn reason is 2 coded.

4. Conclusion

With the strongly development of telecommunication industry, the service providers needs more knowledge of the subscriber. In today's very hard competitive environment, holding of existing customers has become an enormous challenge. Churn analysis is used to predict of customer behaviours that are most likely to change provided service and to compose special marketing tools for them.

Churn prediction in customer relationship management is critical issues in telecommunication industry. For the purpose of competitive in this industry, service providers must be able to predict probability of churners and take proactive approach to retain existing customers. In this research, we use WEKA data mining algorithms for purpose of different classification techniques and clustering for churn prediction.

This study is inspected determinants of customer churn in the Turkish telecommunications industry service market using a sample of have 498.357 actual and churn customer data. The efficiency and the performance is compared of Naive Bayes and Bayes net from Bayes Classifier, Multilayer Perceptron from functions, JRip, PART, OneR from Rules Classifier and Random Forest and J48 from Trees.

The goal of the paper is define and explain the related factors in active and churn prediction modelling. On the other hand this study aims to cluster the application data with data mining simple k means clustered technique. Classified and clustered that uses for churn prediction in Data Mining.

Acknowledgements

The dataset were obtained one of the biggest telecommunication companies in Kayseri, Turkey. The authors wish to thanks University of Administrative and Economy Science, Nuh Naci Yazgan University, Kayseri/TURKEY.

References

[1] Hung S.Y., Yen D.C. and Wang H.Y ,(2006). Applying Data Mining To Telecom Churn Management.. Expert Systems with Applications, vol. 31, pp.515–524.

[2] Kirui C., Hong L., Wilson C., Kirui H. (2013). Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining. IJCS, 10.

[3] Mattison R. (2005). The Telco Churn Management Handbook, Oakwood Hills, Illinois: XiT Press.

[4] Wei C. P. and Chiu I.-T. (2002). Turning Telecommunication Calls Details To Churn Prediction: A Data Mining Approach, Expert Systems with Applications, vol. 31, pp. 103-112.

[5] Huang B., Kechadi M. T. and Buckley B. (2012). Customer Churn Prediction In Telecommunications, Expert Systems with Applications, vol. 39, p. 1414–1425.

[6] Basiri, Taghiyareh and Moshiri (2010). A Hybrid Approach to Predict Churn. Services Computing Conference (APSCC), 2010 IEEE Asia-Pacific, pp.485 – 491.

[7] Saradh and Palshikar (2011). Employees churn prediction. Expert Systems with Applications, vol. 38, pp. 1999-2006.

[8] Geppert K. (2002). Customers churn management. KPMG International, A SWISS Association.

[9] Edelstein H. (2000). Building Profitable Customer Relationships with Data Mining, Two Crows Corporations. [Online].Available: https://books.google.com.tr/