ISSN:2229- 6247

Mustafa Gersil | International Journal of Business Management and Economic Research(IJBMER), Vol 7(2),2016, 560-567

Data Mining Practice in the Scope of CRM: Churn Management in the Telecommunication Sector

Dr. Mustafa Gersil
Celal Bayar University,
Business Administration Dept, Manisa, TURKEY

Abstract

In the intense competitive environment businesses are aware that they need to use information effectively and efficiently to be one step ahead of their competitors. To gain information, it is necessary to know the customer closely, collect and store the date that is left behind at each shopping and use it in a systematic way.

The business, by means of Data Mining, can create models to explain customer behaviors and make assumptions for the future and by the virtue of Customer Relationship Management strategies stemmed from these assumptions are developed to strengthen customer relations, decrease the loss of customers and to gain new customers.

The principle aim of this work is to unveil the information that is hidden in the data warehouse of the businesses by using Data Mining techniques and transform them into strategies within the scope of Customer Relationship Management. In the practical part, by using the data collected between the period of March 2008- August 2008 of one of the leading firms in the telecommunication sector in Turkey the customers which are inclined to terminate their subscription to the business were identified. The practical study has been carried out with the SPSS Clementine program and Logistic regression analysis and Decision Tree algorithms have been used during the process. The details of the practice have been stated step-by-step, essential results have been reached and recommendations to keep customers have been developed. (Due to request of the Institution, data have been published after a certain time period.)

Keywords: Data Mining, Data Warehouse, Churn Management

1. INTRODUCTION

Data warehouse is a wide range data storage that has been created by the combination of customer information collected from different departments of the business in a background system to be utilized later. The “customer information system” that has been created by the systematic record of the information to the database is referred as a data warehouse. (Swift, 2001, p.71)

In other words, data warehouse is topic based, consolidated, tame based, permanent data collection used during the decision making process. (Agrawal and Srikant, 1994, p.5)

Data mining has varying definitions based on the perspective, area of interest and opinion of the person who defines it. Some of the definitions of the data mining found in the literature are provided below (Friedman, 1997, p.1-2):

Data mining: Is an important process in the definition of significant, potentially useful and understandable patterns in the data (Fayyad),

It is the process of extracting unknown, understandable and efficient information from big databases (Zekulin),
It is the discovery and analysis procedures on data stacks for the identification of significant rules and patterns (Berry and Linoff),

It is a cluster of methods used during the information discovery process to differentiate relationships and patterns that were not known among data (Ferruzza),

Data mining consists of extract, transform, and load transaction data onto the data warehouse system, store and manage the data in a multidimensional database system, by using application software analyse the data, provide data access to business analysts and information technology professionals, present the data in a useful format, like a graph or table (Saurkar, A., Bhujade, V., Bhagat, P., 2014)

Different types of data mining tools are available in the marketplace, each with their own strengths and weaknesses. Internal auditors need to be aware of the different kinds of data mining tools available and recommend the purchase of a tool that matches the organization's current detective needs (Patodi, S., Jain, M., Negi, T., 2015)

2. THE PRINCIPLE IMPLEMENTATION AREAS OF DATA MINING

Modern data-mining applications, often called “big-data” analysis, require us to manage immense amounts of data quickly (Rajaraman, Leskovec, 2014, p.36).

Data mining has started to be used in new and different fields every other day. The areas that are most frequently used can be gathered under few headings.

Data Mining in CRM: Considering the fact that the cost of gaining a new customer is much higher than keeping an existing customer and the cost of regaining a customer is much more than acquiring new ones, businesses must be customer oriented and they need to hold on to their existing customers. (http://www.spss.com.tr). The path to customer satisfaction goes through businesses’ better understanding of their customers and increasing their capacity to think more like their customers. By this means identifying customer inclinations is an essential point for the success of the companies. Businesses create assumption models on the future behavior of the customers based on their past behaviors, habits, demographic characteristics. With this aim the customer data coming from different channels is combined in a single environment and leads to significant information that was not known previously. Through this information, the profit of the potential customers who are to become active customers, how long they are to stay as active customers and possibly when they will terminate their subscription to the business and buy product/service from competing business is estimated. (Rygielski, Wang and Yen, 2002, p.494).

Retail Sector: Retail sector is a critical sector, which rapidly develops, and changes that directly interacts with the customer, where competition is very high, customers purchase behaviors change swiftly. By the sales terminal and coding systems in the retail sector, data is collected systematically; connection with the customer is built through store cards and credit cards. This data collected in the retail sector is of vital importance for gaining advantage in the highly competitive environment and thus for the success of the company. With data mining methods hidden information within this data is discovered and the discovered information has critical importance for corporate competition and success. (http://www.spss.com.tr).

Banking and Finance Sector: Through data mining banks can calculate the credit risks of their customers and identify which ones have a high credit risk, as well as estimate which customers will deliver pay back in time. By looking into the characteristics of those who skip their credit card payment, who deliver it with delay or who do not pay at all, they can spot other possible people who might face the same situation. (Özmen, 2001, p.3).

3. PRACTICE

The data set used in this practice is randomly selected from the İzmir, Manisa, Aydın, Muğla, Denizli, Afyon, Kütahya and Uşak provinces of Aegean Region from one of the biggest institutions active in the telecommunication sector in Turkey.

It is created from the records of the months March, April, May, June, July and august from 92.217 subscribers. Upon the request of the institution the data has been released certain period after the practice.

The data belongs to home subscriptions of real persons, which are not legal entities. It is known that 20.871 of the subscribers cancelled their membership between the period of November-December and 71.346 still continues with their subscription.

The subscribers used in the modeling are not all selected among those who were existing subscribers of the business on 01.03.2009. One part has subscribed to the business in March and before, other part during the April, May or June period. Since the telecommunication sector is one where there are rapid customer changers and hence is not stable, having a model where subscribers come to the business in different periods will be more realistic. Another point that needs to be considered that the subscribers that have come in the period of July or August have not been included in the module since the time period is too short.

In the practice after the model that has been created between 01.03.2009 and 31.08.2009 with the customer data, leaving a two-month gap period has been found suitable. The aim here was to create sufficient time for proactive marketing activities. Proactive marketing is the action taken by the business for the subscribers, which have the high possibility to terminate their subscription to the business the business according to the developed module in the forecasted period. For example a change in the tariff which will enable them to talk longer for less, free of cost within city minutes or a gift phone with the commitment to stay at the business for a certain time period are examples of proactive marketing activities.

From the data warehouse of the business, customer data for 6 months have been collected and restructured with Microsoft MS SQL Server 2008 and later on to gather new information that is significant, unknown, previously unrecognized and which might bring added value to the business is used for data mining methods with the Spss Clementine program.
Many firms and individuals prefer Spss Clementine program with its easily understood graphs, strong functional sets, and visual interface. Spss Clementine can be used in many fields from credit card fraud to customer behavior analysis.
A standard procedure for data mining is the point in question. This standard process shall be analyzed in the scope of CRISP-DM methodology. In the upcoming sections, how this process is processed for the data set used in this practice is explained.

3.1. Purpose of the work

Losing their existing customers is one of the biggest problems of today's businesses. In a market where there is intense competition, in sectors where competitors present a similar product and service, customers frequently prefer other competitors.
If the businesses can identify the information of which customers will terminate their subscription to the business the business from the data before the customers actually terminate their subscription to the business, then special measures for those customers can be taken and the rate of customer loss can be decreased.
The purpose in this work is to conduct a customer analysis (churn analysis) to identify the customers who have waived from the products and services that the firm subject to this practice presents.
The aim was to analyze the data with the utilization of Logistic Regression analysis for the subscribers of "active" houses where at least one phone call has been made (call from the house line) and the Decision Tree Technique for "inactive" houses where the subscriber has made no phone calls (no call from the house line).

3.2. Understanding the data

The first step of understanding the data is acquiring it. For this the subscribers info recorded in different tables at the firms database, the properties of the phone line, information in regards to the use of the lines have been collected from the related tables.
The data need to be known and analyzed closely. However one point should not be forgotten during the analysis of the data and that is not to alter it. In this scope it has been analyzed if there are any wrong, missing or repeating records in the data set.
The "Quality" tab of "Data Audit" node of the Spss Clementine program has been used to make the necessary analysis on the variables.

3.3. Preparation of the data

In the preparation phase of the data, the problems identified at the prior stages need to be corrected, needs to be suitable for modeling. Developing a model with incorrect, missing or inconsistent data set will not be a correct model.
In the data interpretation stage the missing values in the Age variable has been extracted from the practice data set. In addition records of records of those under 18 and above 90 of age that are considered as extremes were not included in the data set.

3.4. Modeling

In order to find the best-suited model for the described problem many different models need to be setup and tried. For this reason data preparation and model setting is a renewing stage until the best-suited model is found.
In this work, persons who have made at least one call during their subscription are considered active, and those who have made no calls are consider to be inactive and separate models have been developed for the two groups.

3.4.1. Modeling for active home subscribers

While developing a model for the active home subscribers Logistic Regression Analysis has been used. In the data set prepared for the practice in the case where the subscriber terminate their subscription to the business the business the depended variable (churn) shall be 1, in the case the subscriber stays the variable shall have a double categorized 0. As the demonstration of the double category data is more simple and appropriate, Logistic Regression Analysis is chosen in the practice.
With the different variables in the data set different models have been tested and the model with the highest accuracy has been selected. Figure 1 illustrates different models tested on Spss Clementine for active house subscribers.

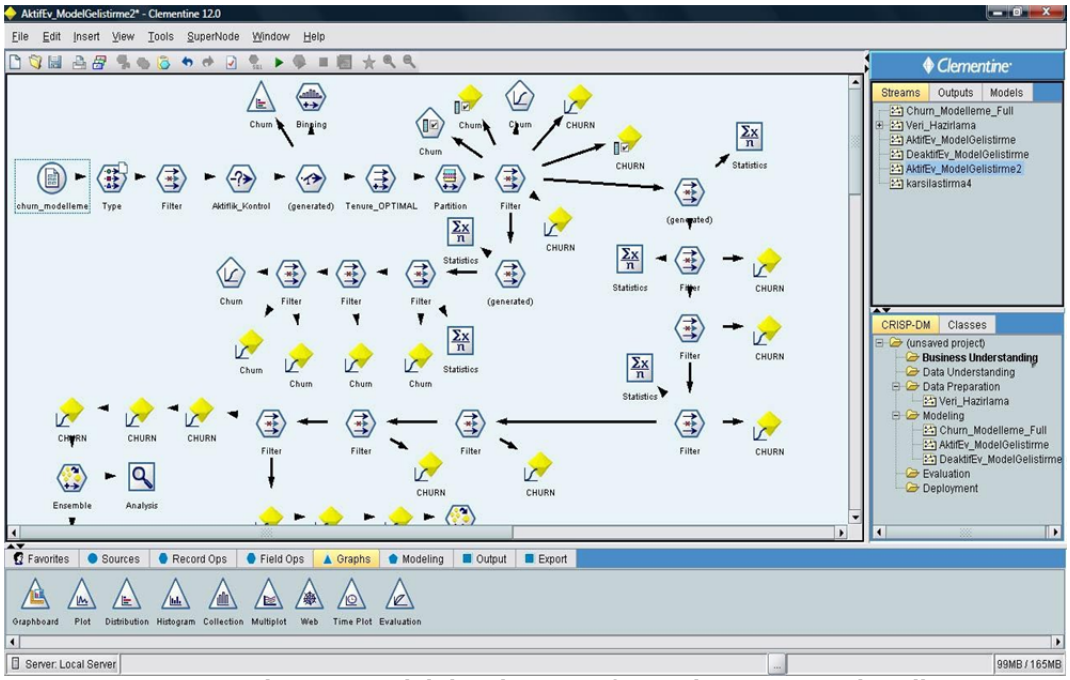

Figure 1 Model development for Active House Subscribers

With the „Statistics" node in the Spss Clementine program the correlation among variables have been analyzed and if there is a high correlation among independent variables, the target variable and the variable with lower relation has not been included in the modeling. To state it otherwise variables that have an effect on the dependent variable however which do not have a relation among themselves have been preferred.
In table 1 B parameters and Wald statistics, degree of freedom, prevention levels and Odds values related to these parameters have been indicated.

Table 1 Results of Logistic Regression

Churn(a)		B	Std. Error	Wald	df	Sig.	Exp(B)	95,0% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
1	Intercept	-4,455	,077	3309,956	1	,000			
	Recency_Out	,005	,001	51,308	1	,000	1,005	1,004	1,007
	MEAN_Arama_Sayisi_LOC	-,003	,000	60,322	1	,000	,997	,997	,998
	W_MEAN_Arama_Yapilmayan_Gun_sayisi	,030	,002	178,962	1	,000	1,031	1,026	1,035
	Gecikmeli_Zamaninda_Orani	,078	,005	247,770	1	,000	1,082	1,071	1,092
	[Adsl=0]	1,468	,038	1503,441	1	,000	4,340	4,029	4,674
	[Adsl=1]	0(b)	.	.	0	.	.	.	.
	[Int=0]	,778	,029	744,387	1	,000	2,177	2,059	2,303
	[Int=1]	0(b)	.	.	0	.	.	.	.
	[Detay=0]	1,185	,028	1745,339	1	,000	3,270	3,094	3,457
	[Detay=1]	0(b)	.	.	0	.	.	.	.
	[Tenure_OPTIMAL=1]	2,594	,091	818,716	1	,000	13,388	11,208	15,991
	[Tenure_OPTIMAL=2]	2,085	,057	1337,213	1	,000	8,044	7,193	8,995
	[Tenure_OPTIMAL=3]	1,962	,052	1442,273	1	,000	7,110	6,426	7,868
	[Tenure_OPTIMAL=4]	1,186	,050	573,067	1	,000	3,274	2,971	3,608
	[Tenure_OPTIMAL=5]	,630	,052	145,089	1	,000	1,877	1,695	2,080
	[Tenure_OPTIMAL=6]	,357	,060	35,367	1	,000	1,429	1,270	1,608
	[Tenure_OPTIMAL=7]	0(b)	.	.	0	.	.	.	.

3.4.2. Modeling for inactive house subscribers

Within the period of 6 months, subscribers who have never used their house phone line are considered to be inactive subscribers. For example, subscribers who have a phone line connected to their summerhouse but who only use it during the summer season or those who have a phone line for the Internet connection but do not have the phone are among the inactive subscribers. In this model, variables such as the usage fee of these subscribers, number of calls made, their average, weighted average or normalized average are not considered.

For inactive subscribers the decision tree technique is found appropriate and among the decision tree algorithms the Quest Decision Tree Algorithm has been chosen.
To find the most suitable model, many different models with the use of different variables have been tested and with the assumed best model a decision tree has been created. Figure 2 illustrates Spss Clementine Stream made to create the decision tree.

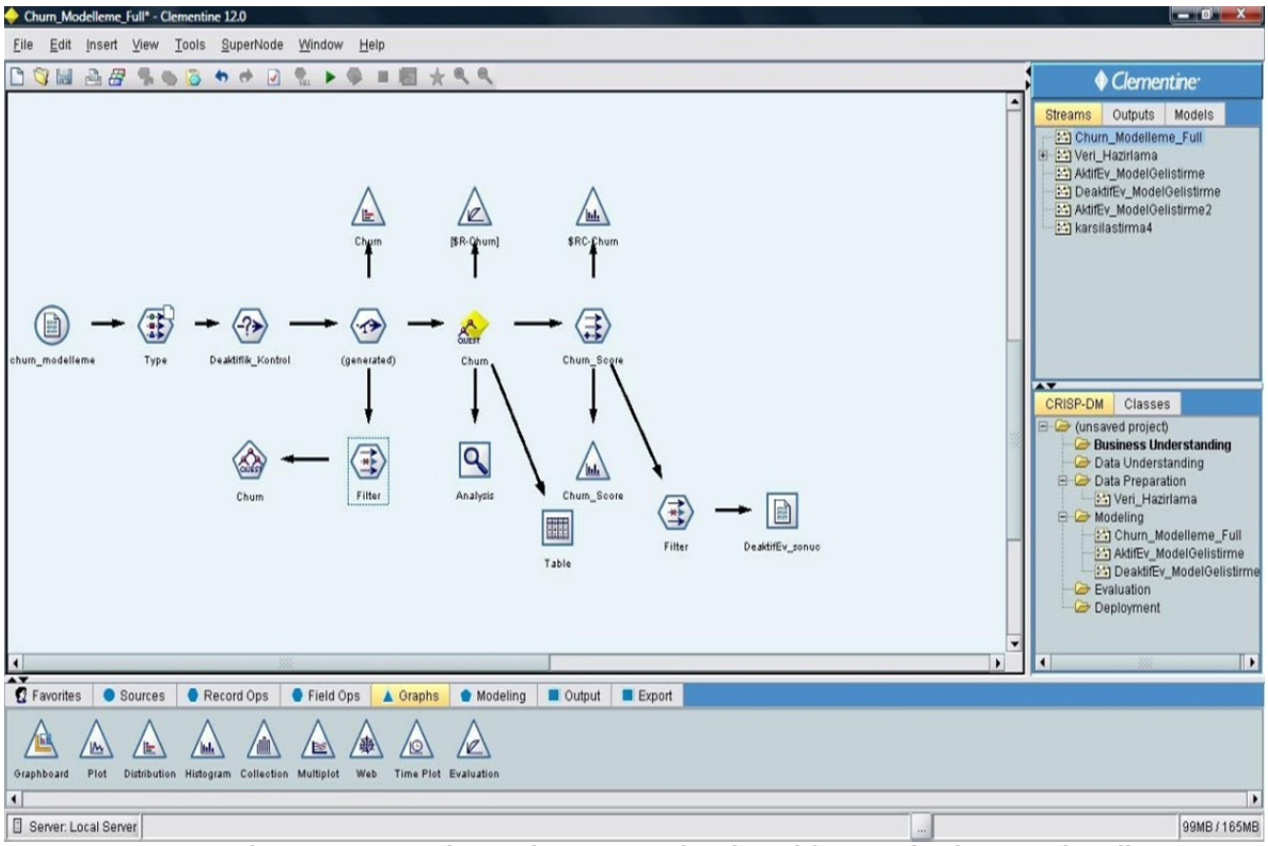

Figure 2: Spss Clementine Stream developed for Inactive home subscribers

3.5 Assessment

The aim in the assessment step is to analyze the results of the conducted data mining and review the conducted procedures. To do this, it is observed if the necessary quality in regards to the significance of the model, accuracy rate is sustained. Since there are two models that are two models, the assessment stage is covered by two headings being the active and inactive stage.

3.5.1. Assessment for active house subscribers

The situation of 19.330 subscribers who left the business is assessed and according to this; 94% of the subscribers who left did not have ADSL, 81% did not get detailed bill, 80% did not open their line for external calls and it is seen that a significant percent of 41%' subscription is less than 135 days. The developed model foresees that the subscribers which do not have ADSL, that do not get detailed bill, whose line is closed for external calls and whose subscription period is not that long have a tendency to terminate their subscription to the business. The revealed results support the model.
When assessed according to the number of calls made; within city calls average for the subscribers who left is 28, for the subscribers who stayed at the business is 56 days. Under these circumstances we can say that the average number of inner city calls are influential in their decision to stay at the business and as the model foresees as this number increases the subscribers continue to stay at the business.
Also the reasons behind the termination of the subscription and the number of subscribers who left the business because of these reasons are analyzed. It has been observed that out of 19.330 subscribers the majority terminated their subscription because they no longer required the services it provided. Another important reason behind the termination is that the communication fee was found expensive.

3.5.2. Assessment for inactive house subscribers

In order to assess the quest decision tree model developed for inactive house member the accuracy rate and gain that the model provides is analyzed.

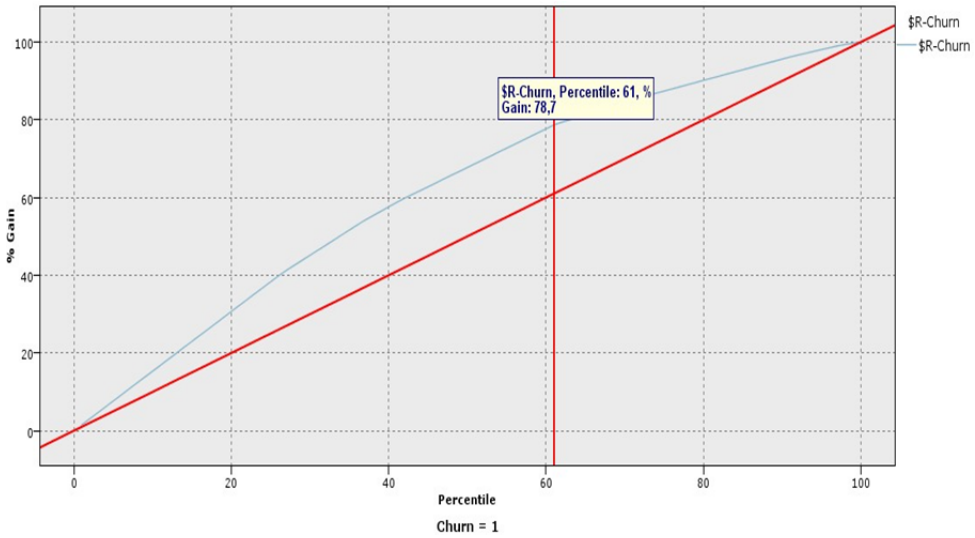

Figure 3 Gain of the model

In figure 3 you may find the gain graphic of the module. According to this when reached 61% which is the highest possibility of termination that the module indicates, in reality the number of subscribers who go with the termination reaches to 78,7%. In this case if 61% part is set target for proactive marketing campaigns to decreases the customer loss, reaching 78,7% of the subscribers before they terminate their subscription to the business and making the necessary campaign announcements and offers will be possible.
According to the data of 1.541 inactive subscribers that left the business, 90% of the inactive subscribers did not have ADSL, 81% do not receive detailed bill and 96% have their line open to external calls. According to the developed model, having ADSL, detailed bill, having a line closed for external calls is effective in terms of keeping the subscribers. While the average of the variable for delayed payment of the bill is 3,5 for the subscribers who have terminated their membership, it is 1,2 for the continuing subscribers. When the rate of delayed payment of bills is higher than the bills paid on time the subscribers are more inclined to terminate their membership. In this case, we can say that the assumptions that the model develops and the results revealed are overlapping.
The reasons behind the termination of subscription are similar in inactive house memberships with the active house memberships. It is observed that out of the 1908 subscribers, 1431 terminated their subscription because there was no need for the services provided.

3.6. Realization

The model that is developed and validated can be a direct application or can be a sub-section of another application. The developed models can be directly used for risk analysis, credit assessment, fraud identification as well as being embedded into applications that allow the automatic order when the predicted inventory levels are bellow the reorder level.
The results acquired in this work can be used in the marketing activities of the business. When the business conducts campaigns on products and services, by prioritizing subscribers that have a high rate possibility of terminating their subscription according to the models the customer loss of the business can be decelerated.

4. Results And Solution Recommendations

In this work, data mining and customer relationship management that gains more importance and becomes more widespread everyday is analyzed in detail, and the data belonging to a firm in the telecommunication sector is analyzed with the data mining technique.

The aim of the work is to cover data mining techniques with CRISP-DM process, conduct the necessary analysis, and extract significant information from data. In this scope, by developing models that present customers who have chosen to terminate their membership in the firm which I of topic to this work, various campaigns and marketing strategies to keep the customers within the scope of CRM have been suggested.
The telecommunication sector is a dynamic sector where there is high circulation. If the profile of the customers, which have an inclination to terminate their subscription, can be identified in advance, special campaigns designed for the target group can be developed and customers can be kept in the business.

Without differentiating active and inactive subscribers, for all the subscribers;

It has been observed that ADSL membership has a positive effect on subscribers loyalty to the business. ADSL is a strong weapon of the firm. In order to receive ADSL services the subscribers first receive phone line services. However if sole ADSL service starts, it is speculated that the rate of termination particularly for inactive subscribers might rise. Plus subscribers who terminate their phone line can receive Internet services from other firms. To prevent this both phone and ADSL services need to be attractive for subscribers. In order to do this the business with is current infrastructure can improve its ADSL service by increasing its bandwidth without changing its price. Faster Internet with the same price will attract subscribers who do not use this service and will be a important factor for the existing subscribers loyalty. In addition, multimedia packages such as free music, video might increase the ADSL use. Furthermore the subscribers need to be informed about other services that they can receive together with the phone line. For example sending short messages to other phones with their phone, visual communication or wakeup calls can be some of the additional services that can attract subscribers. By this means, in addition to communication, which is the main purpose of the phone, these types of additional services might increase the utilization.

It has also been observed that customer who do not receive detailed bills are inclined to terminate their subscription. Since detailed bills show each item in detailed, it allows the subscriber to control the bill sum. In other words the subscriber can come up with conclusions on how to pay less in the following period.
With this in mind, free of charge detailed bills can be sent to subscribers who are inclined to terminate their subscription and who do not receive it in the current state.

Increase in the number of delayed bill payments increases the risk of termination as well. The reason why the bills are not timely paid is due to the fact that the subscriber forgets to deliver it or it can also be financial reasons. The bills that are paid with delay have a daily interest and this leads to a figure that which is much more than the initial one. An unhappy subscriber is inclined to terminate the subscription To avoid the bills that are delayed because they are forgotten, the business can have agreements with the banks and try to convince their subscribers for automatic payment order. This can be supported by various promotions that the bank will also carry out as well. To avoid delays caused by financial reasons, the bill and the payment due date can be customized for the customer after a meeting and the existing debt can be split into installments.

Active subscribers;

Those whose phone is closed to international calls are more inclined to terminate their subscription when compared to those who do have their phones open to international calls. If the business international call tariffs are better than the competitors, this leads the subscribers who benefit from the international call service to stay subscribed longer. People who receive their international call service from different channels can be reached through visual or written media and the price advantages of the business can be delivered.

As the time of the subscriber with the firm increases the possibility of staying loyal also increases. With this it can be said that the new subscribers have a higher rate of terminating their subscription. In order to increase the time of subscription the business can have long-term promotional options. These options may include providing a gift phone to the subscribers with their commitment to stay subscribed for a certain period of time.

As the subscribers' intercity calls numbers increases, it is seen that their risk of terminating decreases. This is an indicator that the business is more preferred in intercity calls in comparison to their competitors. In order to encourage subscribers with low intercity call rates, additional free call minutes in the rate of the inter-city calls they have made can be provided or other tariffs which provide more inter-city call minutes for a lower price can be offered.

Increase in the days with no calls trigger the termination of subscription. This result can be explained, as subscribers do not want to pay for the services that they do not use. For the subscribers to use their phones in a regular way free call time periods within certain hours can be offered.

As the days between the last call made and the time when the model was developed increases, it is observed that the subscribers are inclined to terminate their membership. To sate it otherwise, when the time between calls increase the subscribers are more inclined to terminate their membership. To lead subscribers to make more phone calls, with offering tariffs suitable to different profiles he subscribers can be persuaded to stay. For instance tariffs with discount for different professional groups can be offered.

Inactive subscribers;

It has been observed that the subscribers who have blocked their phones for call are more inclined to stay loyal to the business in comparison to the subscribers who have active phone lines. Since inactive subscribers are people who do not make calls, the reason why they have blocked their phones for external calls can be interpreted by the fact that they use their subscription service for other purposes. For this reason so long as the subscribers who have blocked their phones lines use their subscription for other purposes they will continue to stay subscribed. It is observed that the subscribers who do not use their phone line and who have not blocked it for external calls are inclined to terminate their subscription. For these subscribers to become active subscribers call hours with discount or encouraging campaigns such as “make calls for 30minutes and 10 minutes are from us” can be offered. In addition related campaigns on other services provided by the business can be developed and offered to this group of subscribers. For example, if they have connected a phone line for ADSL services than the phone line will be cut when the ADSL is no longer used. To prevent this for the subscribers who have children going to school, a free of charge Internet education support tool can be provided. In this case the subscriber will continue to use ADSL for the child’s education and therefore shall not terminate the phone line

The information collected in this work and together with the recommendations; will be able to decrease the termination of subscription and also increase the profit margin of the business by improving the existing relationship with the customers, increasing the phone call traffic and also increasing the demand for the other services that the business offers.

REFERENCES

Agrawal, R., Srikant, R., Fast Algorithms for Mining Association Rules, Proceedings of the 20th VLDB Conference Santiago, 1994, p.5

Berry M. J. A. ve Linoff G. S., Mastering Data Mining: The Art and Science of Customer Relationship Management, John Wiley & Sons, NewYork, 2000

Ekinci, T., Veri Madenciligi ve Telekomünikasyon Sektöründe Uygulaması, Yüksek Lisans Tezi, Celal Bayar Üniversitesi, Sosyal Bilimler Enstitüsü, 2009

Friedman, J. H., Data Mining and Statistics: What"s the Connection?, Stanford University, Departmant Of Statistics and Stanford Linear Accelerator Center, 1997, p.1-2

Özmen, S., Is Hayatı Veri Madenciliği ile istatistik Uygulamalarını Yeniden Kesfediyor, V.Ulusal Ekonometri Ve Istatistik Sempozyumu, Eylül 2001, p.3

Patodi, S., Jain, M., Negi, T., Data Mining: The Massive Data Set, “International Journal of Computer Science and Information Technologies”, vol 6, 2015

Rajaraman, A., Leskovec, J., Ullman, J., Mining of Massive Datasets, Stanford University, 2014, p.36

Rygielski, C., Wang, J. ve Yen, D.C., Data Mining Techniques for Customer Relationship Management, Technology in Society, Volume 24, Issue 4, 2002, p.494

Saurkar, A., Bhujade, V., Bhagat, P., A Review Paper on Various Data Mining Techniques, “International Journal of Advanced Research in Computer Science and Software Engineering”, Vol.4, Issue 4, 2014

SPSS Inc., Clementine 12.0 Algorithms Guide, 2007

Swift, R. S., Accelerating Customer Relationships, Prentice Hall PTR A Pearson Education Company, Upper Saddle River, 2001, p.71