Churn Prediction in Telecommunication Industry Using Rough Set Approach

Adnan Amin[1], Saeed Shehzad[2], Changez Khan[1], Imtiaz Ali[1], and Sajid Anwar[1,*]

[1] Institute of Management Sciences Peshawar, Zip Code: 25000, Pakistan
{geoamins,imtiazkmu,sajidanwar.2k}@gmail.com
[2] City University of Science and Technology, Peshawar, Zip Code: 25000, Pakistan
saeedshehzad@gmail.com

Abstract. The Customer churn is a crucial activity in rapidly growing and mature competitive telecommunication sector and is one of the greatest importance for a project manager. Due to the high cost of acquiring new customers, customer churn prediction has emerged as an indispensable part of telecom sectors' strategic decision making and planning process. It is important to forecast customer churn behavior in order to retain those customers that will churn or possible may churn. This study is another attempt which makes use of rough set theory, a rule-based decision making technique, to extract rules for churn prediction. Experiments were performed to explore the performance of four different algorithms (Exhaustive, Genetic, Covering, and LEM2). It is observed that rough set classification based on genetic algorithm, rules generation yields most suitable performance out of the four rules generation algorithms. Moreover, by applying the proposed technique on publicly available dataset, the results show that the proposed technique can fully predict all those customers that will churn or possibly may churn and also provides useful information to strategic decision makers as well.

Keywords: Churn Prediction, Rough Set Theory, Classification.

1 Introduction

Customer churn is one of the mounting issues of today's rapidly growing and competitive telecom sector. The focus of the telecom sector has shifted from acquiring new customer to retaining existing customers because of the associated high cost [1]. The retention of existing customers also leads to improved sales and reduced marketing cost as compared to new customers. These facts have ultimately resulted in customer churn, prediction activity to be an indispensable part of telecom sector's strategic decision making and planning process. Customer retention is one of the main objectives of CRM (customer relationship management). Its importance has led to the development of various tools that support some important tasks in predictive modelling and classification.

[*] Corresponding author.

D. Camacho et al. (eds.), *New Trends in Computational Collective Intelligence*,
Studies in Computational Intelligence 572, DOI: 10.1007/978-3-319-10774-5_8

In recent decades, the organizations are increasingly focusing on long term relationships with their customers and observing a customer's behavior from time to time using various applied knowledge discovery (KDD) techniques [2], [3], [4], [5] to extract hidden relationships between the entities and attributes in a flood of data bank. These facts have attracted many companies to invest in CRM to maintain customers' information. Customer centric approach is very common, particularly in telecommunication sector to predict the customers' behavior based on historical data stored in CRM. Data maintained in such CRM systems can be converted into meaningful information to consider the mounting issue of customer churn to identify the customer's churn activities before the customers are lost which increase the customer strength [6].

Although customer churn prediction modeling has been widely studied in various domains such as financial services, social network services, telecommunication, airlines, online gaming and banking [7]. However, rough set theory application has not been widely studied in customer churn prediction in telecommunication sector. Therefore, this study is approaching to explore the powerful applications of rough set theory for churn prediction in telecommunication sector by constructing more appropriate predictive classifiers that forecast churn prediction based on accumulated knowledge.

The rest of the paper is organized as follows; the next section presents customer churn and related prediction modelling. The preliminary study about rough set theory is explored in section 3 and evaluation methods described in section 4. The evaluation setup and experiments are discussed in section 5 followed by results and comparison in section 6. The paper is concluded in last section 7.

2 Customer Churn and Churn Prediction Modeling

2.1 Customer Churn

Customer churn— Shifting from one service provider to next competitor in the market. It is a key challenge in high competitive markets, which is highly observed in telecommunication sector [5]. Literature reveals the following three types of customer churns [8];

- Active churner (Volunteer): Those customers who want to quit the contract and move to the next provider.
- Passive churner (Non-Volunteer): When the company discontinues the service to a customer.
- Rotational churner (Silent): Those customers who discontinued the contract without prior knowledge of both parties (customer-company).

The first two types of churns can be predicted easily with the help of traditional approaches in term of the Boolean class value, but the third type of churn may exist which is difficult to predict because there may have such type of customers who may possibly churns in near future. It should be the goal of the decision maker and marketers to decrease the churn ratio because it is a well-known phenomenon that existing customers are the most valuable assets for companies as compare to acquiring new one [1].

2.2 Churn Prediction Modelling

Churn prediction has received a tremendous focus from both types of researchers (academia and industry) and have proposed numerous studies for churn prediction in various domains such as financial service [9], Banking Sector [10], [11] Complaint & Repair Services [12], Credit Card Accounts [6], [13], Games [14], [15], Airline Sector [16], Social Networks [17], [18] and Insurance company [19], [20]. Most of the studies have been strongly focused on a few specific factors (e.g: Customer satisfaction& dissatisfaction, loyalty, social influence) relating to customer churn instead of scientifically and empirically investigation and testing of prediction model which encompassing relationships between different constructs such as important churn related variables, switching reasons, service usage, costs and behavior. For instance, in a study [11] classified the customers switching of service with eight general categories. In subsequent work, indicated that the call quality factor is highly influenced on customer churn in the proprietary dataset [21]. Furthermore, surveys based study [22] used a small sample of customer data that may undermine the threat validity and reliability of outcomes.

The literature shows that various machine learning techniques for churn prediction in the telecom industry has been used such as neural network [2],[8],[23],[24], decision table [23],[24],[25], regression analysis [24], SVM [26], [23], Naïve Bayes and Bayesian network [25], evolutionary approach [27] and neuro-fuzzy [28] but conflicts arise when studying results and comparing the conclusion of few of these published works in the area of churn prediction because it is also observed that most of the studies have only evaluated a limited number of traditional machine learning techniques small sample's size [22] and some of them used proprietary dataset. Therefore, the most important problem of which classification technique could use to approach the customer churn prediction in a more appropriate fashion, still remains an open research problem [3]. However, SVM is one of the state-of-the-art technique for classification due to its ability of model nonlinearities but the main drawback is noticed that it generates black-box model [29]. Furthermore, in both studies [8] and [30] stated that artificial neural networks can outperform as compared to other conventional machine learning algorithms.

A benchmarking and empirical study is proposed with the aims to produce further contribution in term of not only achieving suitable accuracy and performance but also extraction of decision rules from hidden existing patterns, based on these rules decision maker can easily adopt new retention policies and improve the overall performance of the organization. In this study, we used rough set theory which has many advantages such as [31]; (1) Finds minimal sets of reduction, (2) Mathematical power to extract hidden patterns from small to large datasets, (3) A straightforward generation and easy interpretation of decision rules. The primarily study about rough set theory is explained in the next section.

3 The Rough Set Theory

Rough Set theory was originally proposed by Pawlak [32] in 1982. A rough set theory has a precise concept of lower and upper approximation and the boundary region. The boundary region separates the lower approximation from upper approximation. Pawlak [33] has defined mathematically rough set approximation and boundaries as; suppose set $X \subseteq U$ and B is an equivalence relation in information system IS= (U, B). Then BL= $\cup$ $\{Y \in U/ B: Y \subseteq X\}$ is lower approximation and exact member of X while BU = $\cup$ $\{Y \in U/ B: Y \cap X \neq \emptyset\}$ is upper approximation which is possibly a member of X. BR = BL - BU is the boundary region.

3.1 Decision Table

The special case of Information system (IS) is known as a decision table [34]. Formally, information system (IS) is IS = (U, A) where U is a non-empty finite set of instances called the universe and A is a non-empty finite set of attributes or properties i.e $A = \{a_1, a_2, a_3, \dots a_n\}$ such that a $:U \rightarrow Va$ for every $a \in A$. A decision system is any information system of the form $S = (U, C \cup \{d\})$, where C are conditional attributes and $d \notin C$ is the decision attribute. The Union of C and {d} are elements of Set A.

3.2 Indiscernibility, Reduct and Core

The information system IS = (U, A), For any subset of attributes $B \subseteq A$ indiscernibility relation IND (B) is defined as follows: If IND (B) = IND (B-{a}) then $a \in B$ is dispensable otherwise indispensable in B while set B can be called independent if all attributes are indiscernible. If $(i, j) \in U \times U$ belong to IND (B) then we can say that i and j are indiscernible by attributes from B. The reduct's notion can be define by indiscernibility relation. If $B \subset A$ and IND (b) = IND (B). The reduction process is finding more important attributes that preserve discernibility relation with the information. If A be the attribute set of universe set U and B is reduced attribute. Therefore, B is equal of sub-set A and the equation for reduced set is composed as; RED (B) $\subset$ A. The core is the intersection of all reducts. Core (B) = Red (B). Where Red (B) is set of all reducts of B.

3.3 Cut and Discretization

By cut a variable $a_i \in$ IS where IS is an information system, such that Value of a_i is an ordered set and it can be denoted as value $c \in Vai$. Actually cut mostly appear in the discretization process as pair which determines a split of interval into two disjoint subintervals. Discretization is a process of grouping the attribute's data based on the calculated cuts and the continuous variables, converting into discrete attributes [6]. There may exists such unseen object which cannot match with the rules or it can increase computational cost that slow down the machine learning process, so cut & discretization methods are used in order to get high quality of classification [34].

3.4 Rules Generation

The decision rules can be constructed by overlaying the reduct sets over the decision table. It can mathematically express as; $(a_{i1} = v_1)$^…..^$(a_{ik}= v_k)$ => $d = v_d$, where $1 \leq i_1 < ... < i_k \leq m$, $v_i \in V_{ai}$; Usually the expression represents as; IF C THEN D where C is set of conditions and D is decision value. To extract the decision rules, the following classification algorithms can be used [34]; (i) *Exhaustive Algorithm (GA):* It takes subsets of features incrementally and then returns the reducts of required one. It needs more concentration because it may lead to extensive computations in case of complex and large decision table. It is based on a Boolean reasoning approach [35] (ii) *Genetic Algorithm:* It is based on order-based GA coupled with heuristic and this evolutionary method is presented by [36], [37]. it is used to reduce the computational cost in large and complex decision table. (iii) *Covering Algorithm:* it is customized implementation of the LEM2 idea and implemented in the RSES covering method. It was introduced by Jerzy Grzymala [38]. (iv) *RSES LEM2 Algorithm:* it is a separate-&-conquer technique paired with lower and upper approximation of rough set theory and it is based on local covering determination of each object from the decision class [38], It is implementation of LEM2 [34].

4 Evaluation Measures

It is nearly impossible to build a perfect classifier or a model that could perfectly characterize all the instances of the test set [39]. To assess the classification results we count the number of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The FN value actually belongs to Positive *P* (e.g. TP + FN = P) but wrongly classified as Negative *N* (e.g. TN + FP = N) while FP value actually part of N but wrongly classified as *P*. The following measures were used for the evaluation of proposed classifiers and approaches.

– *Sensitivity:* It measures the fraction of churn customers who are correctly identified as true churn.

$$\text{Sensitivity (Recall)} = \frac{\mathrm{TP}}{\mathrm{P}} \tag{1}$$

– *Specificity:* It measures the fraction of true non-churns customers, which are correctly identified.

$$\text{Specificity} = \frac{\mathrm{TN}}{N} \Rightarrow 1 - \text{Specificity} = \frac{\mathrm{FP}}{N} \tag{2}$$

– *Precision:* It is characterized the number of correctly predicted churns over the total number of churns predicted by proposed approach. It can formally express as;

$$\text{Precision} = \frac{\mathrm{TP}}{\mathrm{TP + FP}} \tag{3}$$

– *Accuracy:* Overall accuracy of the classifier can calculate by the given formula in equation 4.

$$\text{Acc} = \frac{\mathrm{TP + TN}}{\mathrm{P + N}} \tag{4}$$

- *Misclassification:* Error on the training data is not a good indicator of performance on future data. Different types of errors can be calculated as;

$$\text{Misclassification Error (MisErr)} = 1 - \text{Accuracy} \tag{5}$$

$$\text{Type-I Error} = 1 - \text{Specificity} = \frac{FP}{FP+TN} \tag{6}$$

$$\text{Type-II Error} = 1 - \text{Sensitivity} = \frac{FN}{TP+FN} \tag{7}$$

- *F-Measure:* A composite measure of precision and recall to compute the test's accuracy. It can be interpreted as a weighted average of precision and recall.

$$\text{F- Measure} = 2.\ \frac{\text{Precision. Recall}}{\text{Precision+Recall}} \tag{8}$$

- *Lift:* It tells about the predictive power of classifier arbitrary compared to a random selection by comparing the precision to the overall churn rate in the test set.

$$\text{Lift} = \frac{\text{Precision}}{P\,(P+N)} \tag{9}$$

- *Coverage:* the fraction of the number of cases satisfying both the condition and decision and number of cases satisfying decision only.

$$\text{Cov} = \frac{\text{Number of cases satisfying Condition and Decision}}{\text{Number of cases satisfying Decision}} \tag{10}$$

5 Evaluation Setup

In this section, an analytical environment is setup to perform the proposed technique using rough set explanation system (RSES) [34] and evaluated four different rules generation algorithms (i.e. exhaustive, genetic, covering, LEM2). These experiments carried out to fulfill the objectives of the proposed study and to address the following points also;

- **P1:** *Which features are more indicative for churn prediction in the telecom sector?*
- **P2:** *Which algorithm (Exhaustive, Genetic, Covering, and LEM2) is more appropriate for generating rules sets for rough set classification approach in telecommunication sector?*
- **P3:** *What is the predictive power of the proposed approach on churn prediction in the telecom sector?*
- **P4:** *Can derived rules help the decision makers in strategic decision making and planning process?*

5.1 Data Preparation and Feature Selection

Evaluating data mining approaches on publicly available dataset, has many benefits in term of comparability of results, ranking techniques, evaluating of existing methodologies with new one [40]. In this study, we have used publicly available dataset which can be obtained from URL [41].

Data preparation and feature selection are important steps in the knowledge discovery process, to identify those relevant variables or attributes from the large number of attributes in a dataset which are too relevant and that can reduce the computational cost [36]. The selection of most appropriate attributes from the dataset in hands, was carried out using feature ranking method titled as "Information Gain Attribute Evaluator", using an WEKA toolkit [42]. It evaluates the attributes worth through the information gain measurement procedure as per the class value. It's diverse the selection and ranking of attributes that significantly improves the computational efficiency and classification. After feature ranking, it includes most relevant and ranked attributes in the decision table. The Table 1 describes the selected attributes which are also addressed to P1.

Table 1. List of selected Top ranked attributes reflects the classification performance

Attributes	Description
Int'l Plan	Whether a customer subscribed international plan or not.
VMail Plan	Whether a customer subscribed Voice Mail plan or not.
Day Charges	A continuous variable that holds day time call charges.
Day Mins	No. of minutes that a customer has used in daytime.
CustSer Calls	Total No. of calls made a customer to customer service.
VMail Msg	Indicates number of voice mail messages
Int'l Calls	Total No. of calls that used as international calls.
Int'l Charges	A continuous variable that holds international call charges
Int'l Mins	No. of minutes that used during international calls.
Eve Charges	A continuous variable that holds evening time call charges.
Eve Mins	No. of minutes that a customer has used at evening time.
Churn?	The Class label whether a customer is churn or non-churn.

5.2 Preparation of Decision Table, Cut and Discretization

The preparation of decision table is an important stage of the proposed study of rough set theory based classification. The decision table which consists of objects, conditional attributes and decision attribute are organized in Table 2.

Table 2. Organization of attributes for decision table

Sets	Description
Objects	{ 3333 distinct objects}
Conditional Attributes	{Intl Phan, VMail_Plan, VMailMsg, Day_Mins, Day Charges, Eve_Mins, Eve_Charges, Intl_Mins, Intl_Calls, Intl_Charges, CustServ_Calls}
Decision Attribute	{Churn?}

Cut and discretization is the plausible approach to handle the large data by reducing the dataset horizontally. It is a common approach used in rough set where the variables which contains continuous values is partitioned into a finite number of intervals or groups. The cut and discretization process is carefully performed on the prepared decision table using RSES toolkits. It adds cuts in the subsequent loop one by one for a given attribute. It considers all the objects in decision table at every iteration and generate less number of cuts [34].

5.3 Training and Validation Sets

In data mining, validation is an extremely important step to ensure that the prediction model is not only remembering the instances that were given during the training process but it should also perform the same on unseen new instances. One way to overcome this problem is not to use the entire dataset for classifier's learning process. Some of the data are excluded from the training set as it begins the process to train the classifier. When the training process is completed, then excluded data can be used to validate the performance of the learned classifier on new data. This overall process of model evaluation is called cross validation. The performance of classifiers can be evaluated through several methods which have been discussed in literature [43].

In this study, the holdout cross validation method is used. We divided the data set into two sets: the training set and the test set. After multiple splitting attempts during the experiments, we concluded that the best performance is obtained if the split factor parameter is set to 0.7 using RSES toolkit that randomly splits the data into two disjoint sub-tables. The division of the dataset into training and validation set are performed multiple times with irrespective of class labels and noted the average performance of the classifier to minimize the biases in the data.

5.4 Reduct and Decision Rules Sets Generation

The decision rules can be obtained from the training set by selecting either of the methods (Exhaustive, Genetic, Covering and LEM2). Where Exhaustive and Genetic algorithms are scanning the training set object-by-object and generate rules sets by matching the objects and attributes with reduct while Covering and LEM2 algorithms can induced rules sets without matching with reduct sets using RSES toolkit. In the proposed study, important decision rules are extracted from the training set through one-by-one these four different rules generations algorithms (Exhaustive, Genetic, Covering and LEM2). The decision rules set, specifies the rules in the form of "*if C then D*" where C is a condition and D refers to decision attribute. For example;

```
If Day_Mins=(108.8, 151.05) & Eve_Mins=(207.35, 227.15) &
   CustServ_Calls=(3.5, *)
Then  Churn=(True)
```

Based on these simple and easy interpretable rules, the decision makers can easily understand the flow of customer churn behavior and they can adopt more suitable strategic plan to retain their churn. All the generated decision rules induced from training set are summarized in Table 3.

Table 3. Statistics about rules induced using four methods

Description	Methods for Calculating Rules			
	Exhaustive	*Genetic*	*Covering*	*LEM2*
Total No. of Rules	4184	9468	369	625
# of rules induced that classifying customer as churn	1221	2674	122	160
# of rules induced that classifying customers as non-churn	2963	6715	247	465

6 Results and Discussion

This section reports the evaluation results and discussion on the performances of churn prediction classifiers which is observed during the experiments. The number of churns is much smaller as compared to non-churns customers in the selected dataset, which can provides tough time to churn prediction classifier during the learning process.

6.1 Evaluation of Classifiers' Performance

We have evaluated four different algorithms for rules generation through with rough set based classification approach using RSES toolkit. All these four methods are applied to the same telecom dataset. Table 4 reflects that genetic algorithm performed much better in term of obtaining 98% accuracy, 100% False churn and 98% true churn prediction along with coverage of all instances which is also reported to P2 because the results shows that it is a more appropriate algorithm out of target four. On the other hand, it is also investigated that genetic algorithm have shown more suitable predictive power which is addressed to P3.

Table 4. Evaluation of Four Rules Generation Mehtods through Rough Set Classification Approach

METHODS	TP	FP	FN	TN	COV	PRE	REC	ER	ACC	SPE	FM
Exhaustive	98	39	35	828	1	0.72	0.74	0.074	0.926	0.96	0.726
Genetic	**118**	**19**	**0**	**863**	**1**	**0.86**	**1.00**	**0.019**	**0.981**	**0.98**	**0.925**
Covering	37	41	37	525	0.64	0.47	0.50	0.122	0.878	0.93	0.487
LEM2	52	26	19	571	0.668	0.67	0.73	0.067	0.993	0.96	0.698

6.2 Comparison

Here we first address to P3: What is the predictive power of the proposed approach on churn prediction in the telecom sector? We have presented the sensitivity, specificity, precision, lift, misclassifications, accuracy and F-measure values in Table 4 and then compared the best performed method's prediction performance with other predictive models which are applied to the same dataset. Where A=Neural Network [2], B=Decision Tree [23], C=Neural Network [23], D=SVM [23], E= RBF kernel function of SVM [26], F= Linear kernel function of SVM [26], G= Polynomial kernel function of SVM [26], H=SIG kernel function of SVM [26] and I=Proposed Approach. By comparing the proposed approach with different Churn prediction techniques applied on similar data set, it is clear that the proposed approach performs very well as compared to the previously applied techniques.

Table 5. Predictive Performance of proposed & previous approaches applied to the same dataset

	A	B	C	D	E	F	G	H	I
Sensitivity	81.75	76.47	83.90	83.37	52.12	39.15	59.44	30.49	0.86
Specificity	94.70	79.49	83.50	84.04	94.70	95.32	96.29	93.12	1.00
Precision	66.27	80.60	83.40	84.20	81.90	79.05	80.95	71.43	1.00
Lift	0.0001	0.0002	0.0002	0.0002	0.0007	0.0005	0.0008	0.0004	0.00001
Type-1 Error	5.30	20.51	16.50	15.96	5.30	4.68	3.71	6.88	0.00
Type-2 Error	18.25	23.53	16.10	16.63	47.88	60.85	40.56	69.51	0.019
MisErr	6.76	22.10	16.30	16.30	14.37	22.14	11.44	29.47	0.019
Accuracy	93.2	77.9	83.7	83.7	85.6	77.9	88.6	70.5	0.981
F-Measure	73.2	78.5	83.7	83.8	63.7	52.4	68.5	42.7	0.925

6.3 Features' Analysis

As we have addressed P1 in sub-section 5.1, furthermore, we have examined the features' sensitivity to determine which features are more indicative for churn prediction in the telecom sector to address P1. Fig.1 shows the point of inflection of various variables such as CustServ_Call, Intl_Charges, Eve_Charges, Day_Charges, Day_Mins, Intl_Plan, Eve_Mins and above these points the churn rate decreases and the curve reflect the churn behavior while those features which are below the curve line e.g: Intl_Calls, VMail_Messages and VMail_Plan shows the churn rate constantly decreases. The curve is increasing at an increasing rate up to point of inflections and beyond these points churns behavior increasing at a decreasing rate. This implies that the churn rate is high in those features which are above the curve except Intl_Calls, VMail_Messages and VMail_Plan.

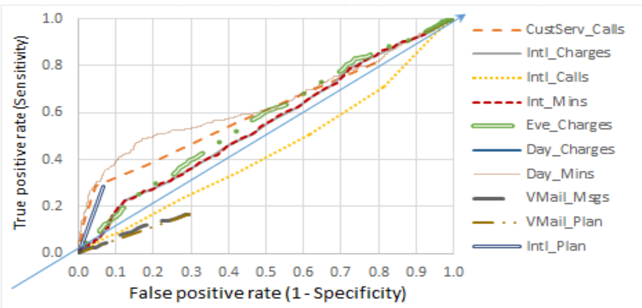

Fig. 1. Reflects sensitivity on y-axis and 1-specificity on x-axis

7 Conclusion

Customer churn is a crucial activity in rapidly growing and competitive telecommunication sector, due to the high cost of acquiring new customers. Churn prediction has emerged as an indispensable part of strategic decision making and planning process. This study is approaching to explore the powerful applications of rough set theory for churn prediction in telecommunication sector by constructing an improved predictive classifiers that can forecast churn prediction based on accumulated knowledge. To evaluate the results of the proposed technique, a benchmarking study is applied and as a result obtained not only more improved churn prediction accuracy, but also other evaluation measures as compared to other state-of-the-art techniques. In this study also investigated the performances of four different algorithms (Exhaustive, Genetic, Covering, and LEM2) of rules generation and extracted useful and easy interpretable decision rules. Based on these rules, the decision maker/manager can easily design more suitable strategic plan to retain the churn.

References

1. Hadden, J., Tiwari, A., Roy, R., Rute, D.: Computer assisted customer churn management: State-of-theart and future trends. IJCOR 10, 2902–2917 (2007)
2. Sharma, A., Kumar, P.: A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services. IJCSA Application 27, 0975–8887 (2011)
3. Wouter, V., David, M., Christophe, M., Bart, B.: Building comprehensible customer churn prediction models with advance rule induction techniques. Expert Systems with Applications 38, 2354–2364 (2011)
4. Kirui, C., Hong, L., Wilson, C., Kirui, H.: Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining. IJCS 10 (2013)
5. Huang, B., Kechadi, M.T., Buckley, B.: Customer churn prediction in telecommunications. Expert Systems with Applications 39, 1414–1425 (2012)
6. Lina, C.S., Gwo-Hshiung, T., Yang Chieh, C.: Combined rough set theory and flow network graph to predict customer churn in credit card accounts. Expert System with Application 38, 8–15 (2011)
7. Yan, L., Wolniewicz, R.H., Dodier, R.: Predicting customer behavior in telecommunications. IEEE Intelligent Systems 2, 50–58 (2004)
8. Lazarov, V., Capota, M.: Churn Prediction, Business Analytics Course. TUM Computer Science (2007)
9. Den Poel, D.V., Lariviere, B.: Customer attrition analysis for financial services using proportional hazard models. European Journal of Operational Research, 196–217 (2004)
10. Chitra, K., Subashini, B.: Customer Retention in Banking Sector using Predictive Data Mining Technique. In: ICIT (2011)
11. Devi, P., Madhavi, S.: Prediction of Churn Behavior of Bank Customers Using Data Mining Tools. Business Intelligence Journal 5 (2012)
12. Tiwari, J., Hadden, A., Roy, R., Ruta, D.: Churn Prediction using Complaints Data. International Journal of Intelligent Technology 13, 158–163 (2006)
13. Lee, K.C., Chung, N.H., Kim, J.K.: A fuzzy cognitive map approach to integrating explicit knowledge and tacit knowledge: Emphasis on the churn analysis of credit card holders. Information Systems Review 11, 113–133 (2001)

14. Kawale, J., Aditya, Srivastava, J.: Churn prediction in MMORPGs: A social influence based approach. IEEE Computational Science and Engineering 4 (2009)
15. Suznjevic, M., Stupar, L., Matijasevic, M.: MMORPG player behavior model based on player action categories. In: 10th Workshop on NSSG. IEEE Press (2011)
16. Liou, J.J.H.: A novel decision rules approach for customer relationship management of the airline market. Journal of Expert Systems with Applications (2008)
17. Oentaryo, R.J., Lim, E.-P., Lo, D., Zhu, F., Prasetyo, P.K.: Collective Churn Prediction in Social Network. In: ASONAM. IEEE/ACM (2012)
18. Guo, L., Tan, E., Chen, S., Zhang, X., Zhao, Y.E.: Analyzing patterns of user content generation in online social networks. In: The 15th ACM SIGKDD, pp. 369–378 (2009)
19. Soeini, R.A., Keyvan, V.R.: Proceedings of Computer Science & Information Technology 30 (2012)
20. Burez, J., Van den Poel, D.: Handling class imbalance in customer churn prediction. Expert Systems with Applications 36, 4626–4636 (2009)
21. Ahn, J.-H., Han, S.P., Lee, Y.-S.: Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. Telecommunications Policy 30, 552–568 (2006)
22. Kim, M.K., Jeong, D.H.: The effects of customer satisfaction and switching barriers on customer loyalty in Korean mobile telecommunication services. Telecom Policy 28, 145–159 (2004)
23. Shaaban, E., Helmy, Y., Khedr, A., Nasr, M.: A Proposed Churn Prediction Model. IJERA 2, 693–697 (2012)
24. Qureshi, S.A., Rehman, A.S., Qamar, A.M., Kamal, A., Rehman, A.: Telecommunication Subscribers' Churn Prediction Model Using Machine Learning. IEEE (2013)
25. Kirui, C., Li, H., Cheruiyot, W., Kirui, H.: Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining. IJCSA 10, 1694–1814 (2013)
26. Innut, B., Churn, G.T.: Prediction in the telecommunications sector using support vector machines. Annals of Oradea University Fascicle of Mgt & Technological Engineering (2013)
27. Au, W.H., Chan, K.C., Yao, X.: A novel evolutionary data mining algorithm with applications to churn prediction. IEEE Trans. 7, 532–545 (2003)
28. Hossein, A., Mostafa, S., Tarokh, M.J.: The Application of Neuro-Fuzzy Classifier for Customer Churn Prediction. Procedia Information Technology & Computer Science 1, 1643–1648 (2012)
29. Farquad, M.A., Vadlamani, R., Raju, B.: Churn Prediction using comprehensible support vector machine: An Analytical CRM application. Elsevier Applied Soft Computing 19, 31–40 (2014)
30. Mozer, M., Wolniewicz, R., Grimes, D., Johnson, E., Kaushansky, H.: Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. IEEE Transactions on Neural Networks 11, 690–696 (2000)
31. Pawlak, Z.: Rough sets, rough relations and rough functions. Fundamenta informaticae 27, 103–108 (1996)
32. Pawlak, Z.: Rough sets. International Journal of Computer and Information Science 5, 341–356 (1982)
33. Zdzislaw, P.: Rough Set: theoretical aspects of reasoning about data. Kluwer Academic Publishers, Dordrecht (1991)
34. Bazan, J., Szczuka, M.S.: The rough set exploration system. In: Peters, J.F., Skowron, A. (eds.) Transactions on Rough Sets III. LNCS, vol. 3400, pp. 37–56. Springer, Heidelberg (2005)

35. Nguyen, H.S., Nguyen, S.H.: Analysis of stulong data by rough set exploration system (RSES). In: Proceedings of the ECML/PKDD Workshop (2003)
36. Bazan, J., Nguyen, H.S., Nguyen, S. H., Synak, P., Wróblewski, J.: Rough Set Algorithms in Classification Problem, pp. 49–88. Physica-Verlag, Heidelberg (2000)
37. Wroblewski, J.: Genetic algorithms in decomposition and classification problem. In: Skowron, A., Polkowski, L. (eds.) Rough Sets in Knowledge Discovery 1, pp. 471–487. Physica Verlag, Heidelberg (1998)
38. Grzymala-Busse, J.: A New Version of the Rule Induction System LERS. Fundamenta Informaticae 31, 27–39 (1997)
39. Burez, J.D., Van den Poel: Handling class imbalance in customer churn prediction. Expert Systems with Applications 36, 4626–4636 (2009)
40. Vandecruys, O., Martens, D., Baesens, B., Mues, C., De Backer, M., Haesen: Mining software repositories for comprehensible software fault prediction models. Journal of Systems and Software 81, 823–839 (2008)
41. Source of Dataset, `http://www.sgi.com/tech/mlc/db/`
42. Holmes, G., Donkin, A., Witten, I.H.: Weka: A machine learning workbench. In: Proceedings of the IEEE Intelligent Information Systems (1994)
43. John, H.: A Customer Profiling Methodology for Churn Prediction. P.hD thesis at Cranfield University (2008)