Journal of Fundamental and Applied Sciences

ISSN 1112-9867

Available online at http://www.jfas.info

Research Article

Special Issue

CHURN CLASSIFICATION MODEL FOR LOCAL TELECOMMUNICATION COMPANY BASED ON ROUGH SET THEORY

M. Makhtar[1,*], S. Nafis[1], M. A. Mohamed[1], M. K. Awang[1], M. N. A. Rahman[1] and M. M. Deris[2]

[1]Faculty of Informatics and Computing, Universiti Sultan ZainalAbidin, Terengganu, Malaysia

[2]Faculty of Computer Science and Information Technology, UniversitiTun Hussein Onn Malaysia, Johor, Malaysia

Published online: 10 November 2017

ABSTRACT

Customer care plays an important role in a company especially in managing churn for Telecommunication Company. Churn is perceived as the behaviour of a customer to leave or to terminate a service. This behaviour causes the loss of profit to companies because acquiring new customer requires higher investment compared to retaining existing ones. Thus, it is necessary to consider an efficient classification model to reduce the rate of churn. Hence, the purpose of this paper is to propose a new classification model based on the Rough Set Theory to classify customer churn. The results of the study show that the proposed Rough Set classification model outperforms the existing models and contributes to significant accuracy improvement.

Keywords: customer churn; classification model; telecommunication industry; data mining; rough set.

Author Correspondence, e-mail: mokhairi@unisza.edu.my

doi: http://dx.doi.org/10.4314/jfas.v9i6s.64

1. INTRODUCTION

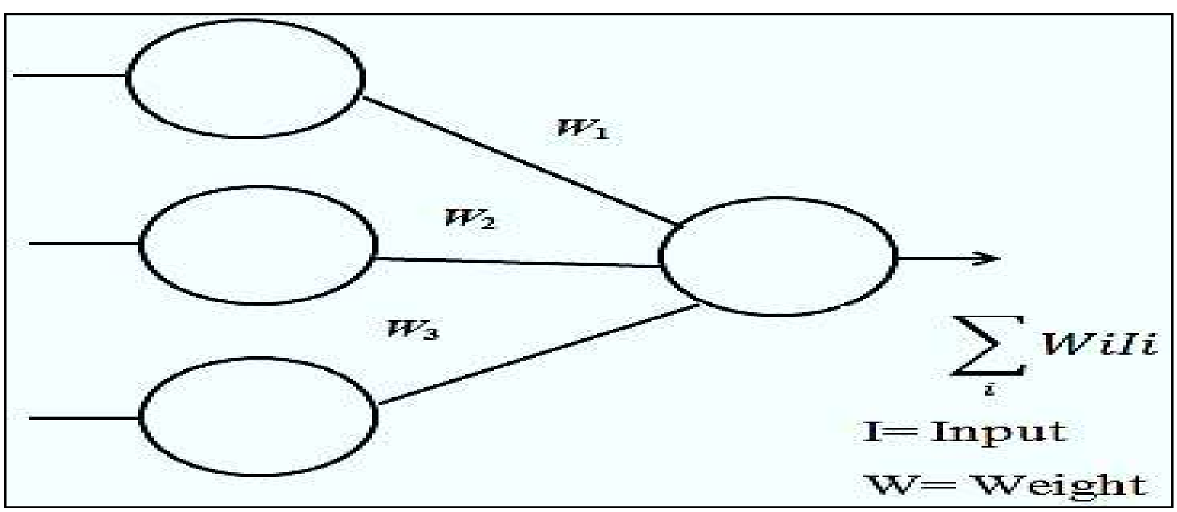
Telecommunication is becoming the number one need in today's environment due to the advancement of computer and network technologies. Due to this phenomenon, the telecommunications industry has become a rapidly growing market. The increasing number of Telecommunication Company has led to intense competition. Thus, customer churn is currently the number one concern for Telecommunication Company. It is important to identify customer churn in the near future since acquiring a new customer is more expensive compared to retaining the existing subscribers [1]. Furthermore, customers can actively exercise their rights to switch from one company to another that satisfies their needs which trigger them to 'churn'.

There are two types of 'churn' such as voluntary churn and involuntary churn [2]. Involuntary churn occurs when customers are disconnected by the service provider for fraud or non-payment. Voluntary churn can be varied and more complex. It can be further classified as incidental churn and deliberate churn. Incidental churn includes customers' financial problems and customers relocating to a new geographical location where the company's service is not available. For deliberate churn, the customer decides to terminate the current service provider due to poor network coverage or poor customer service and chooses to subscribe to the competitor instead. Companies should put more effort to prevent deliberate customer churn since there are many reasons for deliberate churn to occur such as more attractive service packages from competitor, bad network service and matters related to technology. The focus in this paper is to overcome deliberate voluntary churn. By the end of this study, it was found that if firms are fully aware of which segment of the customers poses high risk of churn, the firm can design a treatment program to address the issue. Hence, the key to survive and outperform the industry is to overcome issues that lead to 'churn'.

According to [2], there are two basic approaches to mitigate customer churn namely untargeted and targeted approach. Relying on superior products and mass advertisement to increase brand loyalty and retain customers are categorized as untargeted approach, while in targeted approach, firms will identify customers who are likely to churn and offer direct incentives to avoid that to happen. Targeted approach is divided into two categories namely 1) reactive approach and 2) proactive approach. For reactive approach to occur, the company will wait until customers decide to terminate their subscription. Then, the company offers incentives on the spot. For example, a rebate or cheaper packages, to encourage customer to stay with the current service provider. Meanwhile, for proactive approach, the company tries to identify subscribers who are likely to churn in the near future, then targets these customers with special packages or incentives to keep them from churning. Proactive approach has an advantage in lowering incentive cost. However, this approach will not be effective if the customers are not accurately classified because firms will waste financial resources for wrongly targeted customers. For this reason, customer churn classification should be as accurate as possible. This current research uses the proactive approach to classify deliberate customer churn.

1.1. Related Works

Data mining is a process to extract hidden patterns, relationships and useful information in a bundle of data [3]. Recently, data mining techniques have been used extensively in many fields. For example, they are utilized in mining student's academic [4] and developing innovative applications in agriculture using data mining [5]. Classification is the most vital part in data mining. Generally, classification can be referred to as a process to categorize objects according to the characteristics of the objects. Rapid developments in the classification field allow researchers to develop classification modelling for customer churn using various data mining techniques such as artificial neural network [24-27], decision tree, regression and rough set theory.



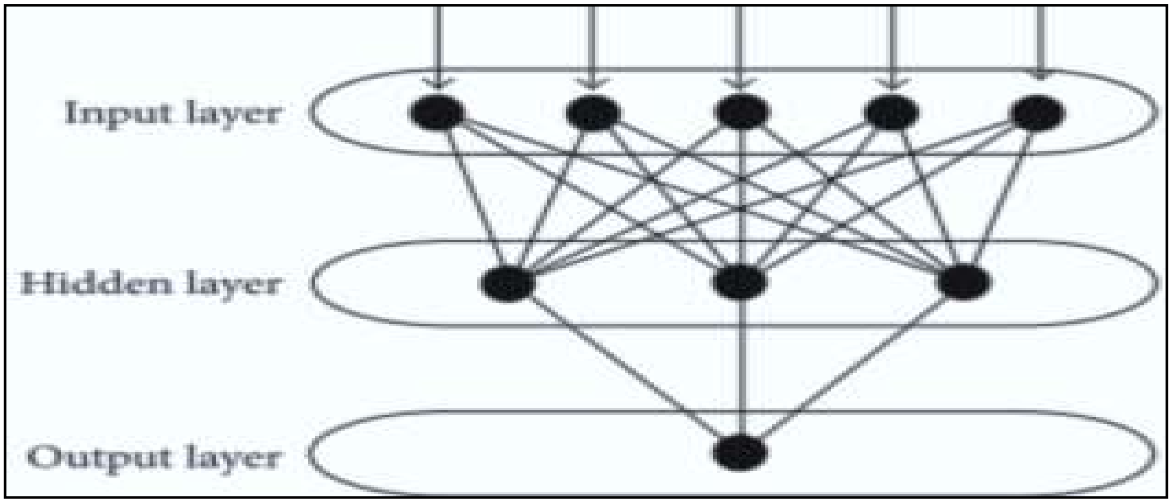


Fig.1. Single layer perceptron neural network and multi-layer perceptron neural network

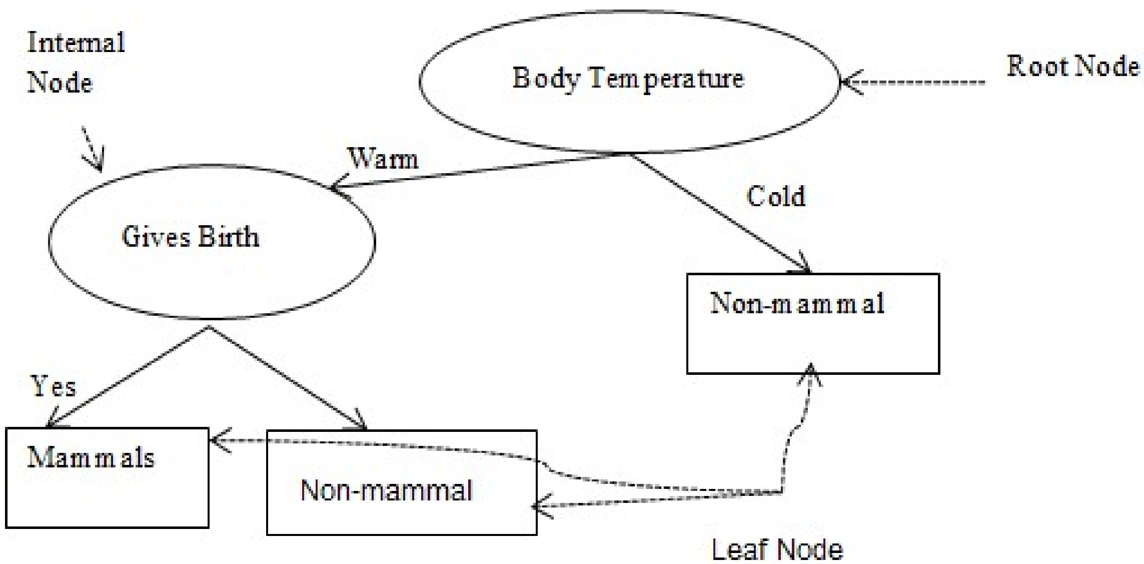


Fig.2. Example of decision tree classification

1.2.Artificial Neural Network (ANN)

Artificial neural network (ANN) can be defined as a model of thought which mimics human brains. In ANN, there are connections between nodes and links. ANN consists of three main elements which are weight, bias and activation function. ANN can be categorized into single layer or multi-layer perceptron (MLP) [23]. Fig. 1 illustrates three elements in ANN for single layer and multi-layer perceptron [6]. ANN approach has advantages in less statistical training however comes with greater computational and proneness to over fitting.

1.3. Decision Tree

Another popular classifier applicable in customer churn risk analysis is the decision tree.

Decision tree is a nonparametric approach for building classification models [7]. In the decision tree, it defines 'nodes' to classify objects. There are three types of nodes: root node, internal node and leaf or terminal node. Leaf or terminal node is assigned to a class label, while non-terminal node (internal and root node) contains attribute test conditions to separate records that have different characteristics. Fig. 2 shows the example of mammal classification problem using the decision tree [8]. According to [9], the decision tree is feasible and effective enough to classify customer churn.

1.4. Regression Analysis

Regression analysis is a statistical classifier approach to investigate relationships between variables. In [10] stated that regression analysis is a good technique for identifying and predicting customer satisfaction. Equation (1) displays a simple linear regression

$$y = a + bx + ei \qquad (1)$$

where y = dependent variables or predicted values, α = constant population value when the value of x is zero, x = value of independent variables, b = constant of independent variables (slope for the population) and e_i = error values or noise or disturbance.

1.5. Rough Set Theory

Rough Set Theory (RST) was introduced by [17] and had attracted many researchers' and practitioners' attentions because this new mathematical approach tackles imperfect and uncertain knowledge. Table 1 summarizes several previous work of customer churn classification based on rough set theory.

Table 1. Previous work for customer churn classification using rough set theory

References	Descriptions
[11]	Utilized RST and back-propagation (BP) neural network
[12]	Classifying customer churn based on historical data by proposing RST-based feature reduction algorithm
[13]	Combined RST and flow network graph to predict customer churn of credit card in Taiwan
[14]	Explored four RST-based reduction algorithms which were Exhaustive, Genetic, Covering and Learning from Example Module (LEM2) to identify the most appropriate algorithm for generating a set of rules for rough set classification

RST guarantees an efficient and feasible algorithm to find hidden patterns and rules in data mining. These hidden patterns and rules can be found through data reduction to make up a minimal set of data. According to [15], RST has advantages in locating minimal datasets. Patterns and rules are in human-readable format which can be easily understood. Furthermore, results obtained can be clearly interpreted and are suitable for parallel processing. In addition, it does not require any preliminary or additional information about data, similar to probability in statistics and grade of membership in the fuzzy set theory. Therefore, implementing the Rough Set Theory for customer churn classification modeling is currently relevant.

2. METHODOLOGY

Rough Set Theory is one of the new generation techniques available for classification. It is an extension of the traditional set approach [16]. Since then, RST has attracted many researchers and practitioners in various fields of science and technology. In classification area, RST has been applied in many real-life applications such as in image segmentation [17], marketing evaluation [18], medical diagnosis [19], stock prices [20] and multimedia data management [21]. The advantages of utilizing RST over other techniques are that it does not need any preliminary or additional information about the data and it offers straightforward interpretation from the obtained results [22]. Moreover, RST is employed to represent imprecise and uncertain information. The elements of the Rough Set Theory consist of indiscernibility relation, lower and upper approximations, as well as attribute reduction.

2.1. Rough Set Theory Approximation Concept

Rough set deals with data analysis in a tabular format called the decision table, which makes up an information system. Fig. 3 illustrates a decision table. Each row in the table represents an object, for instance a case or an event. Meanwhile, each column in the table represents an attribute or feature of the object such as a property or a variable. There are two types of attributes namely condition attribute and decision attribute and each object is assigned with some attribute values.

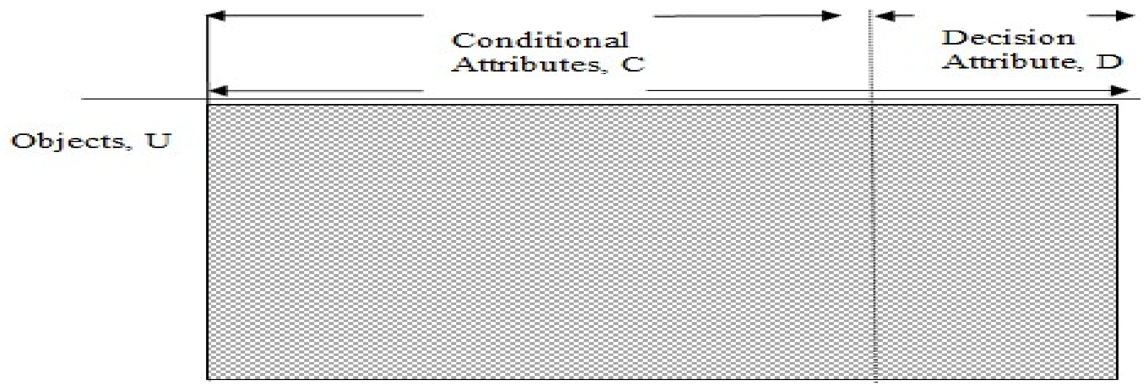


Fig.3. Decision table illustration

Table 2. Example of information table

Obj	Conditional Attribute				
	Sex	**Pckg**	**Monthly Com.**	**Call Plan**	**S. Rental**
O_1	1	1	1	1	1
O_2	1	2	2	1	4
O_3	1	3	3	2	2
O_4	1	4	4	2	2
O_5	0	4	4	2	2
O_6	0	5	5	2	2

Table 2 shows the example of an information table. Then, let IS = (U, A, C, D) be the information system in which U is a non-empty finite set called universe and A is an attribute set. A consists of condition attribute C and decision attribute D such that $A=C \cup D$ and $C \cap D = \emptyset$. D is not necessarily constant on the equivalence class. Therefore, two objects may belong to the same equivalence class but the decision attribute may be varied. For example, if D is inserted into IS in Table 2, it will produce an IS as in Table 3.

Table 3. Example of decision table

Obj	**Condition Attribute**					**Decision Attribute**
	Sex	**Pckg**	**Monthly Com.**	**Call Plan**	**S. Rental**	**Churn?**
O_1	1	1	1	1	1	1
O_2	1	2	2	1	4	1
O_3	1	3	3	2	2	0
O_4	0	4	4	2	2	0
O_5	0	4	4	2	2	1
O_6	0	5	5	2	2	1

From Table 3, it can be observed that O_4 and O_5 are having the same equivalence class with respect to Packages, Monthly Commitment and Call Plan attributes but they are classified differently. This information table can be regarded as inconsistent. As a solution, a condition application of the object is required. Objects O_4 and O_5in Table 3 can also be classified as having indiscernibility relations. The example of indiscernibility relations are:

i. IND (Sex) = $\{\{O_1, O_2, O_3\}, \{O_4, O_5, O_6\}\}$
ii. IND (Call Plan, Service Rental) = $\{\{O_1\}, \{O_2\}, \{O_3, O_4, O_5, O_6\}\}$
iii. IND (Sex, Packages, Monthly Commitment) = $\{\{O_1\}, \{O_2\}, \{O_3\}, \{O_4,O_5\}, \{O_6\}\}$

Based on Equations (3) and (4), the lower approximation and upper approximation of X from Table 3 can be classified as follows:

P(X)-lower= $\{O_1, O, O_3, O_6\}$

P(X)-upper= $\{O_1, O_2, O_3, O_4, O_5, O_6\}$

The boundary as explained in Equation (5) can be defined as $PN_P = \{O_4, O_5\}$.

Indiscernibility relation is the relation between two objects or more where all the values are identical in relation to a subset of a considered attribute. For example, given a subset of attributes,$\alpha \in A$ and $B \subseteq A$, each such subset defines an equivalence relation INDA (B) called an indiscernibility relation that can be defined as follows.

$$(B) = \{(x,x) \in U^2 | \forall \alpha \in B, \alpha(x) = \alpha(x)\} \quad (2)$$

The sets of objects are divided into an equivalence class. In that subset of attributes it will defined a classification process of the universe into sets such that each object in a set cannot be distinguished from other objects in the set using the attributes in B only (refer Equation (2)).

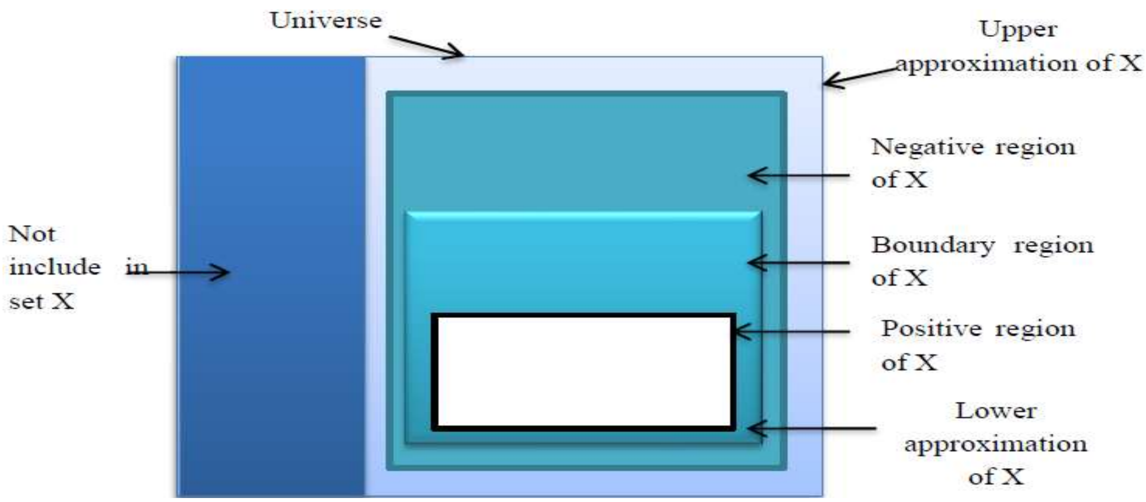


Fig.4.RST approximation concept

The RST concept of border can be expressed by lower and upper approximation. The concept of approximation is required to class the objects based on the equivalence class. Two approximations namely P-lower and the P-upper approximation of X are defined in Equation (3) and (4) respectively where

$$\underline{P(X)} = \{x \in U : [x]P \subseteq X\} \qquad (3)$$

$$\overline{P(X)} = \{x \in U : [x]P \cap X \neq \emptyset\} \qquad (4)$$

Fig. 4 illustrates RST approximation concept. The lower approximation is the set that contains all objects for which the equivalence class corresponds to the object which is the subset of the set. This set contains all objects, which certainly belong to the set X. Meanwhile, upper approximation is the set containing the objects for which the intersection of the object equivalence class and the set is not the empty set. This set contains all objects, which possibly belong to set X. The boundary of X for the given B⊆A and X ⊆U in IS can be defined as

$$PN_P = \overline{P(X)} - \underline{P(X)} \qquad (5)$$

P consists of objects that certainly do not belong to X on the basis of A.

2.2. Attribute Reduction Concept

A reduction set or the so-called ‘reduct’ is a minimal set of attribute after removing redundant and insignificant attributes, but still preserving the original classification. In some cases, not all attributes are required to classify an object. A reduct of A is defined as minimal set of attributes in original classification defined by $B \subseteq A$ such that$\mathrm{IND}_A(B) = \mathrm{IND}_A(A)$. In this example, to discern between the different equivalence classes, only attributes Packages and Call Plan are necessary and the example of reduct is:

$$\mathrm{IND}_A(\{\text{Packages}, \text{Call Plan}\}) = \mathrm{IND}_{(A)}$$

Table 4 shows an example of the decision table after the reduction process, in which attributes Gender, Monthly Commitment and service Rental are dropped. As a result, decision rules in the 4th and 5th rows in Table 4 have the same conditional attributes but different decisions. Thus, rules are considered as inconsistent. Meanwhile, rules in the 1st and 2nd rows are consistent. The RST approximation concept is required in order to handle inconsistency in such decision table.

Table 4. Example of decision table after reduction

Obj	**Decision Attribute**		**Condition Attribute**
	Pckg	**Call Plan**	**Churn?**
O_1	1	1	1
O_2	2	1	1
O_3	3	2	0
O_4	4	2	0
O_5	4	2	1
O_6	5	2	1

From Table 4, the approximation of the decision, D can be defined by constructing a set decision rules. However, decision rules cannot be exactly classified by approximation only. Hence, rules are applied as the implication "if…then..." rules. The rules are constructed as follows:

- Rule 1, if (promotion,1) and (call plan,1) then (churn,1)
- Rule 2, if (promotion,2) and (call plan,1) and then (churn,1)
- Rule 3, if (promotion,5) and (call plan,2) and then (churn,1)
- Rule 4, if (promotion,3) and (call plan,2) and then (churn,0)
- Rule 5, if (promotion,5) and (call plan,2) and then (churn,1)

Thus, it can be concludede that Rule 1, Rule 2 and Rule 3 can be certainly classified as churn. Meanwhile, Rule 4 can be certainly classified as not churn. Lastly, Rule 5 and Rule 6 can be possibly classified as churn and not churn.

3. RESULTS AND DISCUSSION

The data is retrieved from the local Telecommunication Company for some billing period. The dataset contains 21 attributes and 313 objects. However, only 8 significant attributes are retrieved after the attribute selection process. Experiments were conducted using different split factors and reduction methods (Exhaustive Calculation, Genetic Algorithm and Johnsons Algorithm) using Rough Set Technical Analysis Toolkit (ROSETTA).

In order to evaluate the performance of non-RST based and RST-based classifiers, Regression Analysis, J48 (Decision Tree) classifier and Voted Perceptron (Neural Network) classifier were applied to the local telecommunication company dataset using the WEKA software. The same supervised learning approach was utilized in these experiments. For the RST-based classifier, Standard Voting/Tuned (RSES) with Genetic Algorithm reduction method resulting from the previous section was chosen. Table 5 depicts the classification accuracy for different non-RST based and RST-based classifiers.

Table 5. Classification accuracy for different non-RST based and RST-based classifiers

Split Factor	Classifiers with Classification Accuracy (%)			
	L. Regression	J48	V. Perceptron	Rough Set Based Classifier
0.9	19.45	54.80	54.80	90.32
0.8	38.89	64.50	69.40	75.81
0.7	43.85	55.30	73.40	75.53
0.6	44.38	55.30	70.40	76.80
0.5	43.89	55.10	66.00	76.92
0.4	39.13	57.80	68.40	69.52
0.3	31.32	57.80	64.70	71.56
0.2	27.91	56.00	60.40	72.40
0.1	37.12	59.89	54.80	60.85

4. CONCLUSION

This research has attempted to classify customer churn data for a local telecommunication company. The customer churn classification model was successfully developed in this research using RST. In order to yield the best classification accuracy, three reduction attributes were assessed. Genetic Algorithm produced rules with the highest classification accuracies were considered. Several experiments were also performed to test RST-based and non-RST based classifiers. RST-based classifier performed well among the classifiers (Regression Analysis, Decision Tree and Voted Perceptron). After analyzing all the results from the experiment, one decision was made to establish the more appropriate results for this research. All aspects were analyzed to reach the decision that the RST-based classifier was the best classifier to categorize the local telecommunication company's customer churn dataset. Consequently, RST has useful methods that can help to produce better results. In addition, the set of rules produced by the RST model was very informative. These rules can be converted into a predictive system to assist in identifying customer churn for local telecommunication companies.

In conclusion, this research has focused on classifying churners and non-churners in a local telecommunication company. The new classification model obtained is capable of classifying customer churn with the aid of data splitting, feature selection, data discretization, attribute reduction and rule filtering processes.

5. ACKNOWLEDGEMENTS

This work is partially supported by Universiti Sultan ZainalAbidin (UniSZA) and The Ministry of Education, Malaysia (Grant No.FRGS/2/2013/ICT07/UniSZA/02/2).

6. REFERENCES

[1] Aziz A A, Ismail N H, Ahmad F. Mining students' academic performance. Journal of Theoretical and Applied Information Technology, 2013, 53(3):485-495

[2] Rahman M N, Lazim Y M, Mohamed F. Applying rough set theory in multimedia data classification. International Journal of New Computer Architectures and their Applications, 2011, 1(3):683-693

[3] Amin A, Shehzad S, Khan C, Ali I, Anwar S. Churn prediction in telecommunication industry using rough set approach. In D. Camacho, S.W. Kim, & B. Trawiński (Eds.),New trends in computational collective intelligence. Cham: Springer, 2015, pp. 83-95

[4] Bal M. Rough sets theory as symbolic data mining method: An application on complete decision table. Information Science Letters, 2013, 2(1):35-47

[5] Bin L, Peiji S, Juan L. Customer churn prediction based on the decision tree in personal handyphone system service. In International Conference on IEEE Service Systems and Service Management, 2007, pp. 1-5

[6] Cunningham S J, Holmes G. Developing innovative applications in agriculture using data mining. In Southeast Asia Regional Computer Confederation Conference, 1999, pp. 25-29

[7] Hadden J. A customer profiling methodology for churn prediction.PhD thesis, Oxfordshire: Cranfield University, 2008

[8] Ma H, Qin M. Research method of customer churn crisis based on decision tree. In IEEE International Conference on Management and Service Science, 2009, pp. 1-4

[9] Hu D. The study on rough set theory for customers churn. In 4th IEEE International Conference on Wireless Communications, Networking and Mobile Computing, 2008, pp. 1-4

[10] Jiang H, Ruan J. Applying rough set theory to evaluate network marketing performance of China's agricultural products. Journal of Computers, 2010, 5(8):1264-1272

[11] Khoza M, Marwala T. A rough set theory based predictive model for stock prices. In12th International Symposium on Computational Intelligence and Informatics, 2011, pp. 57-62

[12] Backiel A, Baesens B, Claeskens G. Predicting time-to-churn of prepaid mobile telephone customers using social network analysis. Journal of the Operational Research Society, 2016, 67(9):1-11

[13] Lazarov V, Capota M. Churn prediction: Business analyst course. 2007, http://www.vladislav.lazarov.pro/

[14] Lin CS, Tzeng GH, Chin YC. Combined rough set theory and flow network graph to predict customer churn in credit card accounts. Expert Systems with Applications, 2011, 38(1):8-15

[15] Mohammadi G, Tavakkoli-Moghaddam R, Mohammadi M. Hierarchical neural regression models for customer churn prediction. Journal of Engineering, 2013, 2013:1-9

[16]Oseman K, Haris NA, Abu B F. data mining in churn analysis model for telecommunication industry. Journal of Statistical Modeling and Analytics, 2010, 1:19-27

[17]Pawlak Z. Rough sets and data analysis. InIEEE Asian Fuzzy Systems Symposium and Soft Computing in Intelligent Systems and Information Processing, 1996, pp. 1-6

[18]Xie CH, Liu YJ, Chang JY. Medical image segmentation using rough set and local polynomial regression. Multimedia Tools and Applications, 2015, 74(6):1885-1914

[19]Suraj Z. An introduction to rough set theory and its applications. In 1st International Computer Engineering Conference, 2004, pp. 1-39

[20]Tan PN, Steinbach M, Kumar V. Chapter 4-Classification: Basic concepts, decision trees, and model evaluation. In P. N. Tan, M. Steinbach, &V. Kumar (Eds.), Introduction to data mining. Boston: Addison-Wesley, 2006, pp. 145-205

[21]Tripathy B. Chapter 8: Application of rough set based models in medical diagnosis. InS. Dash, & B. Subudhi (Eds.), Handbook of research on computational intelligence applications in bioinformatics. Pennsylvania: IGI Global, 2016, pp. 144-168

[22]Yu R, An X, Jin B, Shi J, Move OA, Liu Y. Particle classification optimization-based BP network for telecommunication customer churn prediction. Neural Computing and Applications,2016, 2016:1-14

[23]Adnan J, Daud NG, Mokhtar AS, Hashim FR, Ahmad S, Rashidi AF, Rizman ZI. Multilayer perceptron based activation function on heart abnormality activity. Journal of Fundamental and Applied Sciences, 2017, 9(3S):417-432

[24]Yassin IM, Jailani R, Ali M, Baharom R, Hassan A, Rizman ZI. Comparison between cascade forward and multi-layer perceptron neural networks for NARX functional electrical stimulation (FES)-based muscle model. International Journal on Advanced Science, Engineering and Information Technology, 2017, 7(1):215-221

[25]Zabidi A, Yassin IM, Hassan HA, Ismail N, Hamzah MM, Rizman ZI, Abidin HZ. DETECTION of asphyxia in infants using deep learning convolutional neural network (CNN) trained on Mel frequency cepstrum coefficient (MFCC) features extracted from cry sounds. Journal of Fundamental and Applied Sciences, 2017, 9(3S):768-778

[26]Hashim FR, Daud NN, Ahmad KA, Adnan J, Rizman ZI. Prediction of rainfall based on weather parameter using artificial neural network. Journal of Fundamental and Applied

Sciences, 2017, 9(3S):493-502

[27]Hashim FR, Adnan J, Ibrahim MM, Ishak MT, Din MF, Daud NG, Rizman ZI. Heart abnormality detection by using artificial neural network. Journal of Fundamental and Applied Sciences, 2017, 9(3S):1-10

How to cite this article:
Makhtar M, Nafis S, Mohamed M A, Awang M K, Rahman M N A, Deris M M. Churn classification model for local telecommunication company based on rough set theory. J. Fundam. Appl. Sci., 2017, *9(6S), 854-868.*