Journal of Theoretical and Applied Information Technology
30th September 2016. Vol.91. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

USING DATA MINING IN TELECOMMUNICATION INDUSTRY: CUSTOMER'S CHURN PREDICTION MODEL

[1]RAYAN MASOUD, [2]TARIG MOHAMED AHMED
[1]MSc Student, University of Khartoum, SUDAN
[2]Assoc. Prof., Department of MIS, Prince Sattam Bin Abdul-Aziz University, KSA
Assoc. Prof., Department of Computer Sciences, University of Khartoum, SUDAN
[2]E-mail: [1]rayanfawzi20@yahoo.com, [2] Tarig_Harbi71@hotmail.com

ABSTRACT

The field of telecommunication is a highly competitive environment. So, telecommunication companies seek to maintain a prominent place in the light of this competitiveness. Also, they seek to acquire new subscribers and retain their subscribers and gain their satisfaction. To ensure that they have a place in this competition, companies must develop their systems and gain knowledge about their subscribers and reasons that may cause them move to their competitors. The using of data mining techniques has become a more common means of acquiring knowledge in a number of different areas. This paper aims to develop a model to predict the churn to rival companies. The model can help companies to know the reasons that may lead to the churn. Three classifiers CN2 (Rules Leaner), Decision Tree and Naïve Bayes were used to build the model. In addition, the dataset was obtained from Columbia University for the purposes of scientific research to develop the proposed model. To evaluate the model performance, F-Measure, classification accuracy, recall and precision were used. The model was implemented using Orange Data Mining Application. The result mentioned that CN2 is the best classification algorithm to build the predicted model.

Keywords: *Telecommunication; data mining; Churn; Classification*

1. INTRODUCTION

Data mining is one of advanced type of analytical tools at this time available; these tools can include statistical models, mathematical algorithms, and machine learning methods (algorithms that increase their performance automatically during experience, such as neural networks or decision trees) but data analysis system that does not deal with large amounts of data [1].

Data mining consists of collecting and managing data and includes analysis and prediction; but this process has limitations to its capability. One limitation is it does not tell the user the value or significance of the patterns and the relationships that were discovered let it to the user. A second limitation is that does not identify a causal relationship that was identified connections between behaviours and/or variables. Because this process to be successful requires skilled technical and logical specialists who can structure the analysis and explain the output that is produced [2].

In the telecommunications industry, huge amounts of data are produced and stored, including: Call data describing the calls that traverse the telecommunication networks, Network data concerning the state of hardware and software components and Customer related data.

Within such tremendous amounts of business critical data, valuable knowledge may be hidden. Indeed, as seen above, it has previously been demonstrated how data mining can be used to discover new patterns and correlations within such sets of data. Applications include telecommunication fraud detection, improving market efficiency, and fault detection and localization. At the same time, a continuous work on customer retention and churn prevention becomes a necessity, because the competition has similar acquisition issues. Retention of the existing users is important since it is 5 up to 7 times cheaper to retain a consumer than to acquire a new one [3].

Two basic groups of churners are voluntary and involuntary churners. Involuntary churners are the customers that Telecommunication Company decides to eliminate from the subscribers list. This class includes customers that are churned for fraud (customers who cheat), non-payment (customers with credit problem), and under-utilization (customers who don't use the phone)[4].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

This paper aims to develop a model to predict the churn to rival companies. The model can help companies to know the reasons that may lead to the churn. Three classifiers CN2 (Rules Leaner), Decision Tree and Naïve Bayes were used to build the model. In addition, the dataset was obtained from Columbia University for the purposes of scientific research to develop the proposed model. To evaluate the model performance, F-Measure, classification accuracy, recall and precision were used. The model was implemented using Orange Data Mining Application. The result mentioned that CN2 is the best classification algorithm to build the predicted model.

2. RELATED WORK

Data mining applications in any field depends on two factors: the available data and the problems facing this field. This section provides Basin Information on the data reserves by Companies of telecommunications [5]. Challenges associated with data and telecommunications are also mining described in this section.

Telecommunications companies reserves data on telephone calls that cross their networks as a Call detail records, which contain descriptive information per phone call. Because of detailed call records maintained for months, billions of detailed calls records available and useful to be mined. Calls details data is useful for applications marketing and fraud detection [7]. Telecommunication companies also maintain large information about Customer, such as invoices information, in addition to information obtained from external parties, Information like credit record. A number of distinct approaches to extract the data has been identified, namely, classification, and association rule learning, clustering, and multi-dimensional scaling (data visualization see). These various techniques can be summarized as follows [8]:

- Techniques of classification supports classify of elements within the data set to predefined classes. Typically in Telecommunication industry the Data mining process involves Customer segmentation, Profiling, Data Preparation and Clustering. Customer segmentation and profiling are equivalent to classification.
- Customer Segmentation is the process of dividing customers into the homogeneous groups according to their common attributes. And customer profiling is the process of describing the customers by their characteristics like age, gender, economic conditions, income, culture etc.

Now, to find the relation between customer profile and segments the Support Vector Machine (VSM) technique can be used. SVM is a Data Mining technique which stands for VSM [9]. SVM estimates segment of a customer by his profile information (age, gender, etc). So if the segment is estimated based on customer profile then we can easily determine usage behavior of that respective customer profile. In today's competition world to compete with other Telecommunication companies, it is very much important to know about your customers and their needs. And hence Customer Segmentation and Profiling is important. Overall the goal of doing all this is to predict the behavior of customer using all the information we have. The profiling is the process done after the segmentation.

Customer segmentation is act of partitioning the customer groups as per their similar characteristics. Customer segmentation helps us to know about the behavior of the customer like what is his usage, calling time, frequency of calling[6] . Customer Profile is built on the basis of his personal information such as age, gender, location. It is done so as to reach the customers and serve them better. Also as per the needs like to offer/build new product the relevant profiles are selected. The basic characteristics of personal information includes- Age and Gender, Location, Economic conditions, Attitude, Lifestyle and Knowledge [10]. Association rule concentrates to discover data elements that co-occur (see events), including the detection of causal relationships. In telecommunications, Association rule learning for instance can be used on a call detail data for identifying couples of customers that calls frequently each other (which in turn can be used to determine the so-called calling circles). Decision tree model classifies the data by arrangement it through the tree to the proper leaf node, which, each leaf node representing classification. Each node represents some attribute Instance. Each leaf node corresponds to one of the possible values for this attributes. Neural networks learns by experience, Popularization of past experiences to new ones, and can be decision-making. It has been used widely in the resolve of many problems in the real world. Two sample rules for classifying a customer as being a business or residential customer are shown below. The rules shown below were generated using a decision tree learner. However, a neural network was also used to predict the probability of a customer being a business or

Journal of Theoretical and Applied Information Technology
30th September 2016. Vol.91. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

residential customer, based solely on the distribution of calls by time of day (i.e., the neural network had 24 inputs, one per hour of the day)[11].

Most telecommunications carriers cluster their mobile customers by billing system data. Billing system data describe customer subscribe, spend and payment behavior. Call detail records describe customer utilization behavior. They have more information to describe customer behavior than billing system data.

Clustering analysis based on call detail records can give more information than other clustering analysis for marketing managements. Defining some new index to describe mobile customer behavior is very useful for forming mobile customer clusters. [12]. Many applications spread out in data mining in industry of telecommunications [13]. However, most of the applications are fall in one of the following three classes:

- Telecommunications Marketing
- Telecommunications Fraud Detection
- Telecommunication Network Fault Isolation and Prediction

As mentioned previously, the inference on how to maintain the customer can be through a variety of different data sources and different techniques. The most important component in any successful system is the use of domain knowledge in the field of telecommunications to draw conclusions from the data [15].

Basically, any solution must (a) be customer-centric, (b) combining information extracted from all available customer data, and (c) be guided by the knowledge domain of telecommunications. That would entail that the system must be able to deal with information in different forms, for illustration, Director's notes in free text, financial, and customer's database.

To meet our objectives and based on all these studies, the model was implemented and evaluated. The classification algorithms were used to generate the model. The following section explains all the steps to build and the evaluation the model in details

3. PROPOSED MOPDEL

The billing system data is the data that describe the customer payment behavior, spend and subscribe. Mobile billing system contains all types of services billing data that the customers will pay every month. Some fields of billing data are shown in table 2. In our study, we select only relative attributes related to research objectives. From business respective, services offered by telecommunication companies beside call and SMS service represent a significant portion of its revenue. So, retaining subscription of customers in these services is major concern of telecommunication companies. Data mining techniques have been broadly used to develop model of churn prediction. In this study, classification tree was applied to develop churn prediction model. After developing model, then it will be evaluated by four evaluation measures including prediction accuracy, precision, recall, and F-measure.

3.1 Instrumentation

As discussed in the previous section, the aim of this research is to classify subscribers based on their Bill data. In order to classify this data using data mining clustering techniques, Orange software will be used. Orange is useful and user friendly and commonly used data exploration tool. In addition, Microsoft office Excel was used in preparing data to be analyzed. In order it used to fill missing values in dataset by random functions.

3.2 Evaluation criteria

At this stage of a project you have built a model (or models) that appears to have high quality, from a data analysis perspective. Before proceeding to final deployment of the model, it is important to evaluate the model more thoroughly, review the steps executed to construct the model, and to be certain that it properly achieves the business objectives. A key objective is to determine whether there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.

3.3 Orange Application

Orange is general-purpose data mining and machine learning tool. It's suitable for different users, beginner data mining users and users who prefer scripting interfaces. Orange was developed by Janez Demšar and Blaž Zupan in 1997. Its Development continued at University of Ljubljana. Orange was designed for first time as C++ Library for machine learning algorithms and related procedures such as data manipulation, data preprocessing and sampling.

3.4 Model Dataset

Telecommunication data faces a large number of strict policies that make access to data extremely difficult. For these reasons, we used the dataset which was obtained from the Center for

Journal of Theoretical and Applied Information Technology
30th September 2016. Vol.91. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Customer Relationship Management at Duke University's Fuqua School of Business. Churn dataset consist of two groups: data group and demo group. Data group consist of six fields:

1. Subscript id: This is the identification of subscriber. All the subscribers in the dataset are consumers. There are no business users.
2. Bill Month: This is the bill month in SAS format.
3. Plan chosen: This variable shows the plan that was chosen by the consumer for that month. It takes 4 values. A value of 1 corresponds to a 200 peak minutes plan, a value of 2 corresponds to a 300 peak minutes plan, a value of 3 corresponds to a 350 peak minutes plan and a value of 4 corresponds to a 500 peak minutes plan. The pricing scheme for the four plans is given in Table 1.
4. Total Minute Qty sum: For each month, this is the total peak minutes that was consumed.
5. Lag Prom Dummy: A dummy variable that records whether a promotion was sent to a customer in the past month.
6. Churn: This is an indicator that takes a value 1 if the consumer defects at the end of the month. It takes a value 0 otherwise.

Table 1 Values of Plan Chosen Attribute

Plan Chosen	No of Minutes	Price of Access	Marginal Price
1	200	30	0.4
2	300	35	0.4
3	350	40	0.4
4	500	50	0.4

Second dataset used in development of churn prediction model is Demo group which consists of six fields:

Churn dataset consist of integrated attributes from both data and demo dataset. Churn dataset attributes illustrated in Table 2.

Table 2 Schema of Churn Dataset

Attributes	Description
Subscript id	To identify Subscriber
Bill Month	Bill of Month in SAS format
Plan chosen	To identify plan of service chosen by subscriber
Total Minute Qty sum	Number of minutes per Month that subscriber uses service
Lag Prom Dummy	To identify whether promotion was sent to subscriber past month
Churn	To record whether subscriber churn after Month
Birth DT	To record Subscriber birth date
Svc-start-DT	Date of start subscription of service
Svc-end-DT	Date of end subscription of service
Zip Code	To identify Subscriber Zip Code
New Cell-Ind	To record whether subscriber is new Cell Phone user

3.5 Experiment

The case company is one of Telecommunication companies that provide services. Dataset under study consist of 3000 subscribers and their information In the following sub-sections we will discuss steps of developing Churn Prediction Model

.

3.6 Data Preprocessing

The aim of preprocessing step is preparing dataset to be suit for purpose of analysis. Churn dataset contains a lot of missing values that will affect the result of discovered model. Missing values in churn dataset filled by random values generated by RANDBETWEEN function in Microsoft Office Excel. Churn dataset consist of two tables: data table and demo table. Demo and data integrated into one table (called churn Table). Churn table consist of eleven attributes, illustrated in Table 2. Some of these attributes unrelated to target of knowledge discovering. Because of this reason, relative attributes selected to be analyzed and remove unrelated attributes. In addition, attributes have different data types Discrete, Continuous and String. Discrete data type means

Journal of Theoretical and Applied Information Technology
30th September 2016. Vol.91. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

that attribute can have finite value (also called Nominal). Continuous mean Numeric data type. String data type means collection of characters. As we say previous, the aim of this research is develop churn prediction model by classification techniques. Classification techniques require discrete attributes. For this, we discretize step done to convert each Continuous attribute to Discrete. Attributes that converted to discrete are Bill-Month and Total Minute Qty sum. Discretization step achieved after attribute selection step using Orange software as shown in Figure 2.

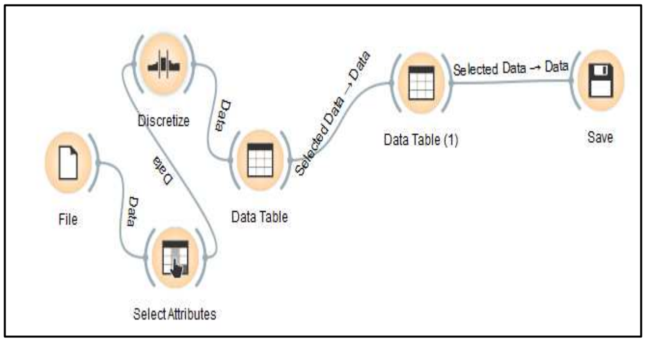

Figure 1. Data preprocessing

Figure 1 illustrates steps of churn dataset preprocessing which done in three steps. First step loading dataset from location where exist. Second step select relative attributes required to discover knowledge from churn dataset. Final step convert continuous attributes to discrete.

Mechanism of Discovering Knowledge

To construct Churn Prediction Model using classification techniques, churn dataset divided into two datasets training and testing dataset. Training dataset represents two-Third of Churn dataset which selected randomly from whole dataset, where testing dataset represents remaining of Churn dataset. Training dataset will used to construct prediction model and testing dataset to evaluate it, show Figure 2.

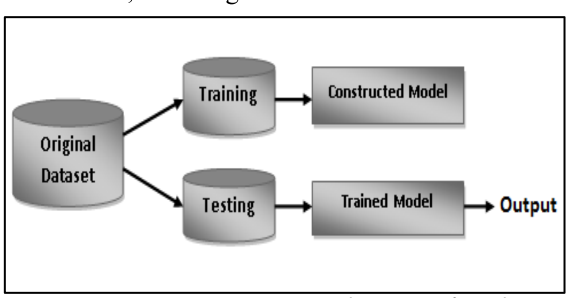

Figure 2: Training and Testing of Prediction Model

In addition to testing dataset we use other dataset to evaluate and validate churn prediction model.

Figure 2 illustrates that original dataset divided randomly into training and testing dataset to construct and evaluate prediction model respectively.

The churn prediction models are based on three CN2, Classification Tree and Naïve Bayes. After dataset divided into training and test dataset, training dataset linked with Data Table widget to show data instances and data table linked with CN2 Rule Learner, Classification Tree Learner and Naive Bayesian Learner. CN2 Rule Learner widget linked with CN2 Rule Viewer to represent discovered knowledge in If-Then Rules. Classification Tree linked with Classification Tree linked with classification Tree Graph to represent knowledge as Tree. Three learners linked with Test learner widget. Test learner do two things, it shows a table with different performance measures of the classifiers, such as classification accuracy and area under ROC. Second, it outputs signal with data which used by another widget to analyze performance classifier such as confusion matrix. Finally, Test learner linked with Confusion matrix.

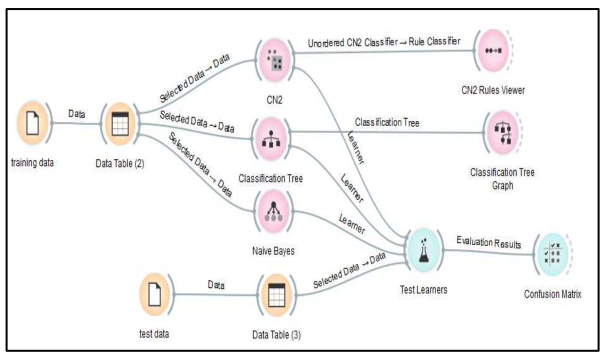

Figure 3: Classification Model

Figure 3 illustrates how to construct classification model. Classification model built using training dataset by classifier learner and evaluated using test dataset by Test learner which output data used to evaluate model.

3.7 Model Evaluation

In this case study, we consider accuracy, precision, recall, F-Measure and Brier as measures to evaluate model churn prediction. These measures calculated using confusion matrix.

Because three classifiers as shown in Figure 4 have equal values of F-measure, Accuracy,

Journal of Theoretical and Applied Information Technology
30th September 2016. Vol.91. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Precision and Recall, Brier used to differentiate between three classifiers. Brier is measure of average deviation between the predicted probabilities of events and the actual events.

Table 3 Performance of Churn Prediction Model

Model	Precision	Recall	Accuracy	F-Measure	Brier
Classification Tree	0.9796	1.000	0.9796	0.9896	0.0400
CN2	0.9796	1.000	0.9796	0.9896	0.0393
Naïve Bayes	0.9796	1.000	0.9796	0.9896	0.0395

1.2.9 Results of CN2

As we see in Figure 4, CN2 viewer output results in form of rules and information about these rules such as rule length, rule quality, Coverage and predicted class. Rule length is number of attributes participate in incident part, rule quality is measure of rule goodness and coverage is number of covered learning examples, predicted class is the class label that predicted by this rule. All these rules could be used to classify the expected churn customers. Each one has the length of number of attributes used by the rule and the percentage of the quality

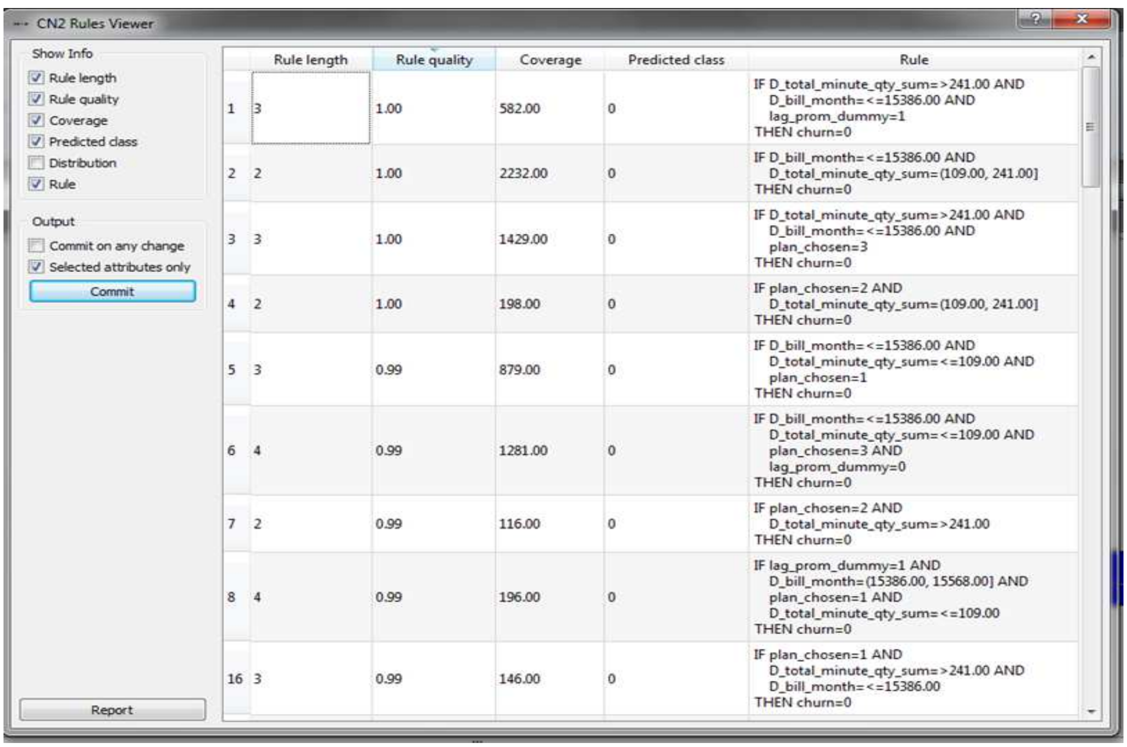

Figure 4: CN2 viewer output

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. CONCLUSION

In telecommunication industry, it is essential for companies that provide services to clients to focus on the most important issues in the management of customer services i.e. churn prediction. Using data mining techniques to improve the performance of companies and help them to predict the occurrence of churn. Many algorithms have been proposed for classification in Customer Churn prediction. We choose four very commonly used algorithms such as CN2, Classification Tree to towards our classification of customers in Telecommunication industry. Among all classifier, our experimental results show that Naïve Bayes achieves highest AUC. In other hand, the worst classification was Classification Tree the algorithm achieves lowest AUC.

From the results of the experiments in table 3, it was found that CN2 was the best way to construct the model followed by the classification Tree and then by Naïve Bayes according to Brier evaluation measure. In addition, experimental results provide some knowledge. For instance, some of the variables that have been selected among the variables related to the subject of the study are the factors that affect the churn such as month's bill and lag-promo-Dummy.

As future direction of this research, the dataset should be improved by adding new attributes to enhance the quality of the model. These attributes must reflect the behavior of the customers in different countries.

REFRENCES:

[1] H. Newton, Newton’s Telecom Dictionary, CMP Books, http://www.cmpbooks.com
[2] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth (1996), Knowledge discovery and data mining: Towards a unifying framework. In: Proceedings of the 2nd ACM international conference on knowledge discovery and data mining (KDD), Portland, OR, pp 82–88
[3] Sheth, Jagdish, and Rajendra Sisodia. The 4 A's of Marketing: Creating Value for Customer, Company and Society. Routledge, 2012.
[4] Kraljević, Goran, and Sven Gotovac. "Modeling data mining applications for prediction of prepaid churn in telecommunication services." AUTOMATIKA: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije 51.3 (2010): 275-283.
[5] Ravisankar, Pediredla, et al. "Detection of financial statement fraud and feature selection using data mining techniques." Decision Support Systems 50.2 (2011): 491-500.G.
[6] Vesanen, Jari, and Mika Raulas. "Building bridges for personalization: a process model for marketing." Journal of Interactive Marketing 20.1 (2006): 5-20.
[7] Privacy Preserving Data Mining in Telecommunication Services) Granmo, O., and VLADIMIR A. Oleshchuk. "Privacy preserving data mining in telecommunication services." TELEKTRONIKK 103.2 (2007): 84.
[8] Customer Segmentation and Customer Profiling for a Mobile Telecommunications Company Based on Usage Behavior, S.M.H Jansen, July 17, 2007.
[9] Kitts, Brendan, Gabor Melli, and Karl Rexer. "Data Mining Case Studies." The First International Workshop on Data Mining Case Studies, 2005 IEEE International Conference on Data Mining, Huston, USA. 2005.
[10] Groth, Robert. Data mining: building competitive advantage. USA;: Prentice Hall PTR, 2000.
[11] Lin, Qining. "Mobile customer clustering analysis based on call detail records." Communications of the IIMA 7.4 (2007): 95-100.
[12] Groth, Robert. Data Mining: A hands on approach. Prentice Hall, 1998.
[13] Sumathi, Sai, and S. N. Sivanandam. Introduction to data mining and its applications. Springer, 2006.
[14] Weiss, Gary M. "Data mining in telecommunications." Data Mining and Knowledge Discovery Handbook. Springer US, 2005. 1189-1201.
[15] Yahia, Moawia Elfaki, and Murtada El-mukashfi El-taher. "A New Approach for Evaluation of Data Mining Techniques." International Journal of Computer Science Issues (IJCSI) 7.5 (2010).