Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCACI-2015 Conference Proceedings

E-Customer Classification using Data Mining for CRM

Sakshi Sivarama Krishna
Associate Professor
KMM Institute of Technology and Sciences
Tirupati, India

Cheruku Sudarsana Reddy
Asst Professor
KMM Institute of Technology and Sciences
Tirupati, India

Harsha
Asst Professor
KMM Institute of Technology and
Sciences Tirupati, India

Abstract: **A significant portion of customer community today is e-customer. Customer relationship management is always inevitable and inevitable in the competitive business situation. In E-business environment there is no lack in data about customer. The only thing of business people to consider is to analyze the data to evaluate the e-customer behavior. Data mining is a good tool to do the same .In this paper we classify the E-business customer to elevate a loyal customer among others .A loyal customer is the backbone of any business to prosper. A classification algorithm is developed and used to classify the E-customer for the purpose of decision making as a part of Customer Relationship Management (CRM).The results are compared with other means of classification. Finally suggestions are proposed to E-business people to improve their CRM.**

I.INTRODUCTION TO E-BUSINESS AND CRM

"CRM is about acquiring and retaining customers, improving customer loyalty, gaining customer insight, and implementing customer-focused strategies. A true customer-centric enterprise helps your company drive new growth, maintain competitive agility, and attain operational excellence" [SAP].CRM is about understanding, anticipating and responding to customers' needs.

E–Business today has exploded many folds due to the invention of new Information and Communication (ICT) technologies and applications. Business models of today are highly depending upon internet and intranet based platforms, use of Web portals and mobile applications [2]. Business data pertaining to E-business has become highly valuable in recent times. Social networking environments, sharing of product data and its reviews in blogs are means for promoting business in new methods. E-business will become a vital competitive strategy that will revolutionize the global economy. Companies are practicing to manage customers' relationships by virtually serving their needs "24 × 7″-24 hours a day, 7 days a week.E-business enables customers to personalize and customize products or services will flourish. Producing, marketing, and distributing products or services online is a cost-effective strategy for business. There is a need to serve online communities with niche interests to build customer loyalty.E-business models provide greater choice for customers and change the traditional economics of supply and demand. Seamless access to the Net from multiple gateways-cable TV, satellite, wireless telephones, and other devices-will greatly expand e-business opportunities. Highly efficient e-business virtual supply chains link manufacturers and producers directly to customers [3]. E-business will reach billion of people and generate trillions of revenues worldwide in near future.

Customer is the vital element in any business situation. It is essential to build healthy relationship with the customer to strengthen the business situation. The e-customer today is having plenty of alternatives to go with to buy the products and services online. It is the duty of the online business to attract and retain a potential customer. To build such loyalty it is needed to establish a good customer relationship management. This is only possible when the e-business organizations have plenty of data regarding the e-customer. A good CRM has to provide the facility for the business to store and manage data they collect on their customers, and products. A better CRM will have the ability to collect the data, convert them to information that enables business improvement.

2. CLASSIFICATION IN DATAMINING AND ITS SIGNIFICANCE IN E-BUSINESS

In Classification: Classification is one of the most useful learning models in data mining. It is aimed at building a modelto predict the customer behaviorsthroughclassifying the database records into a number of predefined classes based on certain criteria. Common tools usedfor classification is neural networks, decision trees and if-then-else rules. The purpose of the Classificationapproach is to predict customer behavior. Classification of Database records based on the defined criteria provides good knowledge.

Classification technique in Data mining provides us a rule set to classify the data [4]. The data to construct the rule set is called training data and the data to be classified using the rule set is called test data. In E-business transactions there is a lot to mine to know the thinking process of the customer in their purchase activity. The

knowledge mined from these transactions provide a way to manage customer relationships [5].

Data mining tools helps CRM by providing the framework, which covers: i) to analyze the business problem ii) to prepare the data requirements iii) to build the suitable model with respect to business problem and, iv) to validate and evaluate the designed model [6].CRM consists of the following dimensions :(1) Customer Identification; (2) Customer Attraction; (3) Customer Retention; (4) Customer Development. These four dimensions can be seen as a closed cycle of the CRM system.

3. ORGANISATION OF THE PAPER

Initially the concepts of E-Business and CRM are highlighted. The purpose of classification as a technique of data mining followed by its significance in E-Business is discussed. In the next section various classification algorithms are discussed. Later the algorithm used in the current work is explained. The discussion on the data set adopted for the purpose followed by the experimental results is presented in the next section. In the next section the strength of the used algorithm is highlighted using comparisons. Finally the significance of the results in E-Business is discussed

4. CLASSIFICATION ALGORITHMS

There are plenty of algorithms in the literature for classifying data.EC 4.5, C 4.5 and CART are some of them in the front line. C 4.5 is considered to be the state of the art algorithm being used and proved its potentiality in many situations. The rest of the algorithms come next in the line of importance. All these algorithms worked base on the statistical function known as Information Gain [7].

Classifiers are commonly used tools in data mining. Classification systems take as input a collection of tuples, each belonging to one of a small number of classes and described by its values for a fixed set of attributes, and output a classifier that can accurately predict the class label to which a new record belongs.C4.5 , a descendant of CLS and ID3 . Like CLS and ID3, C4.5 generates classifiers expressed as decision trees, but it can also construct classifiers in more comprehensible rule set [8]. Decision tree plays an important role in constructing the required rules of classification. This is the simplest approach to represent the structure of the rule management using which the frame work to classify the new arrivals can be managed. It is easy to interpret, needs no parameters, and follows divide and conquer greedy strategy in the construction process. The time complexity of the process is O(log n). Decision tree is the most important and most popular classification technique because it is more interpretable than other state-of-the-art classification techniques, support vector machines and neural networks. Decision tree classifier is generally used as a bench mark before using any other classification technique. It handles both numerical and categorical attributes.A decision tree is built top-down from a root node and follows partitioning the data into sub trees that contain tuples with similar values. ID3 algorithm uses entropy to calculate the similarity of a sample.

If the sample is completely similar, the entropy is zero and if the sample is an equally divided it has entropy of one.

5. PROPOSED ALGORITHM

The decision tree approach as in C 4.5 algorithm is guided our present work in building the rule set to form a frame work to classify the data. The step by step procedure of the process is outlined here.

DECISION_TREE (T)

1. If all the training tuples in the node T have the same class label then
2. set $probability(T) = 1.0$
3. return tree node (T)
4. If (tuples in the node T have more than one class) then
5. Find_Best_Split_Attribute(T)
6. For i ← 1 to datasize[T] do
7. If (split_atribute_value[t_i] <= split_point[T]) then
8. Add tuple t_i to left[T]
9. Else
10. Add tuple t_i to right[T]
11. If(left[T] = NIL or right[T] = NIL) then
12. Create probability distribution of the node T
13. return(T)
14. If (left[T] != NIL and right[T] != NIL) then
15. DECISION_TREE(left[T])
16. DECISION_TREE(right[T])
17. return(T)

6. DATA SET USED FOR EXPERIMENTS

Most promising E-business transactions are linked with bank transactions. The behavior of the customer can be linked with the bank gateway portion from which the customer gone through [8][9].

The online transactions of e-business customers are the dependable source for training as well as the test data in e-business analysis [9][10].For the algorithm developed the online transactional data is adopted with aggregated attributes. The attributes of each record constitutes {loyal_points, transaction_ status, transaction problem, success_rate, transaction_ amount, age, loyalty}. These attributes are the aggregated objects and the attribute values are derived. The class label is loyalty, which gives the potentiality of the customer. To judge whether a customer is loyal or not is based on the aggregated data inference. The portion of the data is given in table 1.

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCACI-2015 Conference Proceedings

loyal_points	trans_status	trans_problem	success_rate	trans_amount	age	loyalty
low	completed	bankgateway	good	high	young	1
low	completed	bankgateway	good	high	old	1
low	completed	bankgateway	good	low	young	0
low	completed	bankgateway	good	low	old	1
low	completed	bankgateway	bad	high	young	0
low	completed	otherproblem	good	high	old	1
medium	completed	customerexit	good	low	young	1
medium	completed	customerexit	good	low	old	1
medium	completed	customerexit	bad	high	young	1
high	completed	bankgateway	good	high	young	1
high	completed	bankgateway	good	high	old	1
high	completed	bankgateway	good	low	young	0
high	completed	bankgateway	good	low	old	1
high	completed	bankgateway	bad	high	young	0
high	notcompleted	customerexit	bad	high	young	0
high	notcompleted	customerexit	bad	high	old	0
high	notcompleted	customerexit	bad	low	young	0
high	notcompleted	customerexit	bad	low	old	0
high	notcompleted	otherproblem	bad	low	young	0
high	notcompleted	otherproblem	bad	low	old	0
high	notcompleted	customerexit	good	high	young	1

Table 1: Sample data

7. EXPERIMENTAL RESULTS

An aggregate set of attributes are obtained by observing large amount of transactions. The data instances are processed to get the decision rules. The following is the resulted tree from the executed process.
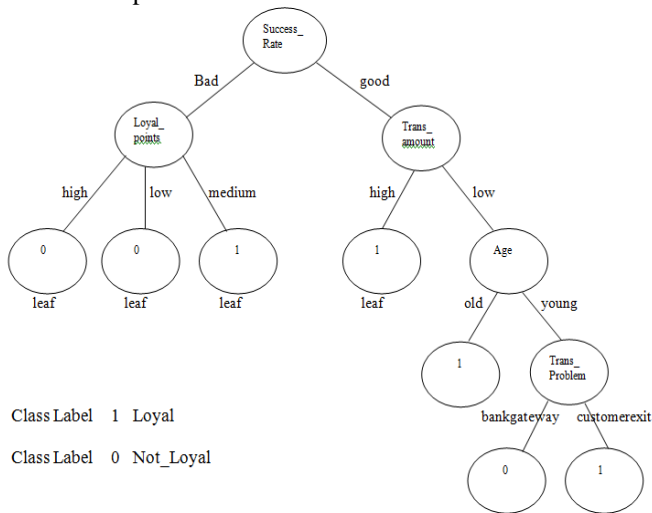

Figure1.Decision tree

The result set provide us a mean to classify new customer information to find whether he/she is a loyal customer for the business or not. Using this information the CRM activity of the business people can be improved..

The factors completeness, effectiveness and maintainability are the key indicators for test code quality. And defect resolution speed. Throughput and productivity are the key indicators for issue handling performance. So, first we have to measure these indicators and try to get relation between them.

8. THE STRENGTH OF THE PROPOSED ALGORITHM

C 4.5 is a frontline algorithm for classification. The successor of this algorithm is followed here to construct decision tree for the selected data set. This algorithm is the mostly proven algorithm for classification. In terms of process time, scalability and utilization of memory, this is the most dependable approach. Since decision trees are better means for rule construction and interpretation of data, and the approach guided by C 4.5 is dependable we adopted the approach to rely upon.

8. SIGNIFICANCE OF THE RESULTS

The proposed algorithm is providing competitive results compared to other algorithms in the field. The algorithm is providing better accuracy. It is scalable with increasing levels of the data. The results obtained are compared with the results found from other means of software like WEKA. The results found proved the strength of the algorithm in terms of scalability, accuracy and process time.

9. CONCLUSION

Classification techniques of data mining are fine means for classifying large business transactions for decision making. Customer Relationship management is a vital task of any business entity today. For retaining the customer it is needed to know the behavior of the customer. The required behavior can be obtained from classification of transactional data obtained from customer interactions in online business. In this work we rely upon classification algorithms to elevate the loyal customer. The technique can be adopted for any business transactional data to assist CRM.

REFERENCES

[1] Arun Kumar Agariya, Deepali Singh "*crm scale development &validation in Indian banking sector*" Journal of Internet Banking andCommerce, April 2012, vol. 17, no. 1.

[2]Kapanen, R. (2004) 'Customer relationship management and service delivery', *International Journal of Services Technology and Management,* Vol. 5, No.1 pp. 42 - 55.

[3]BabitaChopra ,VivekBhambri, ,BalramKrishan "*Implementationof Data Mining Techniques for Strategic CRM Issues*" published in"International Journal of Computer Technology and Applications July-August 2011 Vol 2 (4), 879-883 ".

[4]Jiawei, H. and Kamber, M. (2006) '*Data Mining Concepts and Techniques',* Morgan Kaufmanns, pp.4-27.

[5] Mark Lavender, (2004),"*Maximizing customer relationships andminimizing business risk*", International Journal of Bank Marketing,Vol. 22 Iss: 4 pp. 291 – 296.

[6] MosadZineldin, (2005),"*Quality and customer relationshipmanagement (CRM) as competitive strategy in the Swedish banksindustry*", The TQM Magazine, Vol. 17 Iss: 4 pp. 329 – 344.

[7] XIAOHUA HU "*A Data Mining Approach for Retailing BankCustomer Attrition Analysis*" Applied Intelligence 22, 47–60, 2005Springer Science + Business Media, Inc.

[8]Leo Breiman, Jerome H. Friedman, Richard A. Olshen and Charles J. Stone. Classificationand Regression Trees. Wadsworth & Brooks, 1984.

[9] VivekBhambri "*Application of Data Mining in Banking Sector*"International Journal of Computer Science and Technology Vol. 2,Issue 2, June 2011.

[10] Ogwueleka and Francisca Nonyelum "*Potential Value of DataMining for Customer Relationship Marketing in the BankingIndustry* " Advances in Natural and Applied Sciences, 3(1): 73-78,2009.