CRM Based On Naive-Bayesian Classification

Ruma Panda[1]
Assistant Professor & Research Scholar, CSE Dept.,
Vemana Institute of Technology, Bangalore, Karnataka, India

DR. A. N. Nandakumar[2]
Professor & H.O.D., IS Dept., New Horizon College Of Engineering
Bangalore, Karnataka, India

***Abstract*— CRM is a combination of several components. The companies can process the customer information before beginning the process. Companies can acquire about their customers through their internal customer information or they can purchase data from outside sources. So this paper mainly focused for issuing the beneficial card to get the concession in CRM field using Naive-Bayesian classification. It is very important to classify customers automatically and effectively to get the concession based on their relevant documents in the generation of card.**

Keywords- Data Mining, CRM, Naive-Bayesian Classification.

I. INTRODUCTION

CRM is the midpoint for many industries such as Banking, Retail, Telecommunication and Insurance. CRM takes customer as central point and optimizes the business process to get profit. But in the real world application there are major challenges for building high performance CRM classification model. The main issue in CRM is customer classification and prediction. So that, a company can classifies its customers into predefined groups with same patterns. Data Mining uses sophisticated statistical processing or artificial intelligence algorithms to discover useful patterns and knowledge from the extracted data. So it is very important to know their customer better using prediction models and associations. They will find the valuable information according to the customer's needs by analyzing their raw data. Therefore customer classification and prediction is most important in CRM. Now, here a customer classification and prediction model will be implemented that uses collected information of customers as input to make a prediction for getting discount in the field of CRM by generating the beneficial card. Naive-Bayesian classification has used to implement this.

II. LITERATURE REVIEW

CRM has many applications in the marketing industry. Now a days data mining is a main research area. Data Mining has multiple disciplines, including machine learning, statistics, database systems, information science, visualization and many application domains, has made great progress in the past decade [1].The technological development has enabled new data mining approaches to be applied for finding the best CRM strategies, where data mining can apply a significant role in analyzing customer data[2].Customer Attraction which depends direct marketing[3],Customer identification include target customer analysis and customer segmentation[4], Customer Retention which is based on the satisfaction of the customers[5] and Customer Development includes lifetime value analysis, up/cross selling and market basket analysis[6]. A new feature selection method is proposed to resolve the CRM data set with relevant features by an efficient data mining techniques to improve data quality and feature relevancy after preprocessing [7]. Data Mining strategies and techniques are used to generate the new rules and patterns [8]. Association rules algorithms have used to mine the data to detect the association rules in each time period [9]. CRM principles provide a strategic and tactical focus for identifying and realizing sources of value for the customer and the firm and can guide organizational process [10].A new design scheme of CRM systems based on data mining is presented in [11].

III. METHODOLOGY

Here Naive-Bayesian classification algorithm has used to classify the customers based on their relevant documents to get the discount in the field of CRM by generating the beneficial card.

A. NAIVE-BAYESIAN CLASSIFICATION ALGORITHM

It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods. Bayes theorem provides a way of calculating posterior probability P(c|x) from P(c),P(x) and P(x|c).

1. Let D be the training set of tuples and their associated class labels. Each tuple is represented by an n-dimensional attribute vector, $X=(x_1,x_2,\ldots\ldots x_n)$, depicting n measurement s made on the tuple from n attributes, respectively $A_1,A_2,\ldots A_n$.

2. Suppose that there are m classes, $C_1,C_2,\ldots C_m$. Given a tuple, X, the classifier will predict that X belongs to the class having highest posterior probability, conditioned on X. That is, the Naive Bayesian classifier predicts that tuple X belongs to the class C_i iff

$P(Ci|X) > P(Cj|X)$ for $1 \leq j \leq m, j \neq i$.

The class C_i for which $P(C_i|x)$ is maximized is called the maximum posterior hypothesis. By Bayes' theorem

$$P(ci|x) = \frac{P(x|Ci)P(ci)}{P(x)}$$

Above, P

- $P(c|x)$ is the posterior probability of *class* (c, *target*) given *predictor* (x, *attributes*).
- $P(c)$ is the prior probability of *class*.
- $P(x|c)$ is the likelihood which is the probability of *predictor* given *class*.
- $P(x)$ is the prior probability of *predictor*.

3. As P(X) is constant for all classes, only $P(X|C_i)P(C_i)$ need be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, that is $p(C_1)=P(C_2)=\ldots\ldots..=P(C_m)$,and then maximize P(X|Ci). Otherwise, maximize $P(X|C_i)P(C_i)$. The class prior probabilities may be estimated by $P(C_i)=|C_i,D|/|D|$, where $|C_i,D|$ is the number of training tuples of class C_i in D.

4. Given data sets with many attributes, it would be extremely computationally expensive to compute $P(X|C_i)$. In order to reduce computation in evaluating $P(X|C_i)$, the naive assumption of class conditional independence is made. This presumes that the values of the attributes are conditionally independent of one another, given the class label of the sample, that is, there are no dependence relationships among the attributes. Thus, The probabilities $P(x_1|C_i)$, $P(x_2|C_i)$ …, $P(x_n|C_i)$)i can be estimated from the training samples, where

(a) If A_k is categorical, then $P(X_k|C_i)= S_{ik}/S_i$ where S_{ik} is the number of training sample of class C_i having the value x_k for A_k , and S_i is the number of training samples belonging to C_i.

(b) If A_k is continuous-valued, then the attribute is typically assumed to have a Gaussian distribution so that

$$g(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

5. In order to classify an unknown sample X , $P(X|C_i)P(C_i)$ is evaluated for each class C_i . Sample X is then assigned to the class C_i if and only if $P(X|C_i)P(C_i) > P(X|C_i)P(C_i)j$ for $1 \leq j \leq m, j \neq i$. In other words, it is assigned to the class C_i for which $P(X|C_i)P(C_i)$ is the maximum.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

IV. IMPLEMENTATION

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

Suppose the CRM system hope to increase the customers to use the system by using their CRM beneficial card. There is a large number of valuable customer information in huge amounts of data accumulated by customers which is used to identify customers and provide decision support. We wish to predict the class label of an unknown sample using Naive Bayesian classification, given the training data as Table I.

TABLE I

Sex	Age	Student	Occupation	Disable	Concession
Female	<25	yes	unemployed	No	yes
Male	<25	yes	unemployed	No	yes
Female	25~30	yes	service	No	no
Male	25~30	yes	service	No	no
Female	<25	yes	unemployed	yes	yes
Male	<25	yes	unemployed	yes	yes
Female	31~60	No	service	yes	no
Male	31~60	No	service	yes	no
Female	31~60	No	service	No	no
Male	31~60	No	service	No	no
Female	31~60	No	unemployed	No	no
Male	31~60	No	unemployed	No	no
Female	>60	No	senior	No	yes
Male	>60	No	senior	No	yes
Female	>60	No	senior	yes	yes
Male	>60	No	senior	yes	yes
Female	25~30	No	service	No	no
Male	25~30	No	service	No	no
Female	25~30	No	unemployed	No	no
Male	25~30	No	unemployed	No	no
Female	25~30	No	unemployed	yes	yes
Male	25~30	no	unemployed	yes	yes
Female	25~30	yes	service	yes	yes
Male	25~30	yes	service	yes	yes
Female	31~60	No	unemployed	yes	yes
Male	31~60	No	unemployed	yes	yes

Let C1 correspond to the class card processing with concession =”yes” and C2 correspond to the class card processing with concesion=”no”. The unknown sample we wish to classify is

X=(Sex=”female”,age=”31~60”,student=”no”,occupation=”unemployed”,disable=”no”)

We need to maximize P(X|Ci)P(Ci), for i=1,2.

$P(C_i)$ the prior probability of each class, can be computed based on the training samples:

P(cardprocessing_concession=”yes”)=14/26=0.5385

P(cardprocessing_concession=”no”)=12/26=0.4615

To compute P(X|Ci),for i=1,2, we compute the following conditional probabilities:

P(sex=”female”|cardprocessing_concession=”yes”)= 7/14=0.5

P(sex=”female”| cardprocessing_concession =”no”)=6/12=0.5

P(age=”31~60”|cardprocessing_concession=”yes”)=2/14=0.1423

P(age=”31~60”| cardprocessing_concession =”no”)=6/12=0.5

P(student=”no”|cardprocessing_concession=”yes”)=8/14=0.5714

P(student=”no”|cardprocessing_concession=”no”)=10/12=0.8333

P(occupation=”unemployed”|cardprocessing_concession=”yes”)=8/14=0.5714

P(occupation=”unemployed”|cardprocessing_concession=”no”)=4/12=0.3333

P(disable=”no”|cardprocessing_concession=”yes”)=4/14=0.2857

P(disable=”no”|cardprocessing_concession=”no”)=10/12=0.8333

Using the above probabilities, we obtain

P(X|cardprocessing_concession=”yes”)=0.5*0.1423*0.5714*0.5714*0.2857=.00664

P(X|cardprocessing_concession=”no”)=0.5*0.5*0.8333*0.3333*0.8333=.05786

P(X|cardprocessing_concession=”yes”)P(IDprocessing_concession=”yes”)=.00664*.5385=.0036

P(X|cardprocessing_concession=”no”)P(IDprocessing_concession=”no”)=.05786*.4615=.0267

Therefore, we predict cardprocessing_concession=”no” for sample X. Now the above example illustrates the Bayesian classification algorithm with training data. We can implement the above algorithm using c#.net and SQL database.

V. RESULT ANALYSIS

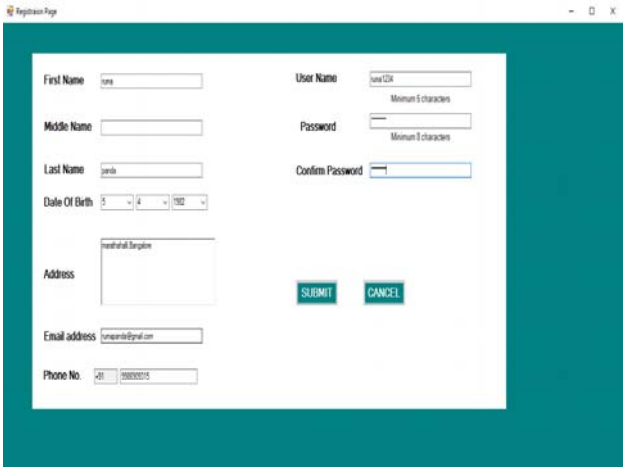

Figure 1: Registration Details

The customer has to register by providing their relevant documents for accessing the system. Figure 1 shows the registration form which one will be filled by new user.

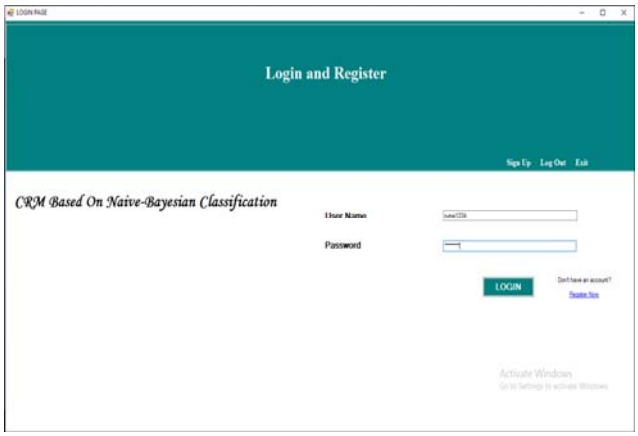

Figure 2: Login Form

After getting registered, the user has to enter the corresponding username and password to login into the system. If the user name and password is not proper then it will display the message. And the user has to give correct user name and password. Figure 2 is showing the login form.

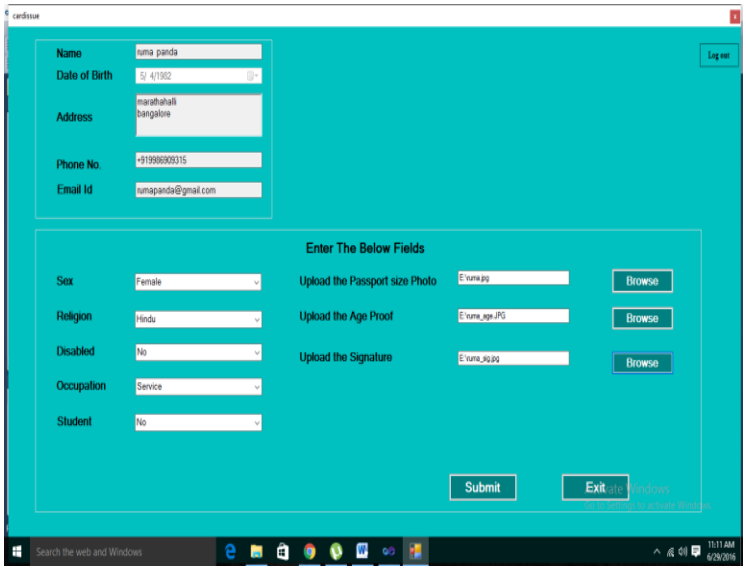

Figure 3: Issuing the Beneficial Card

First time if the user wants to issue a beneficial card then they has to fill the above form by using their relevant documents. User has to upload their passport photo, age proof document and signature. Disable person has to upload the disable proof. This beneficial card will help to get the concession in different field of CRM based on the table1. The concession amount will be dependent on the company people.

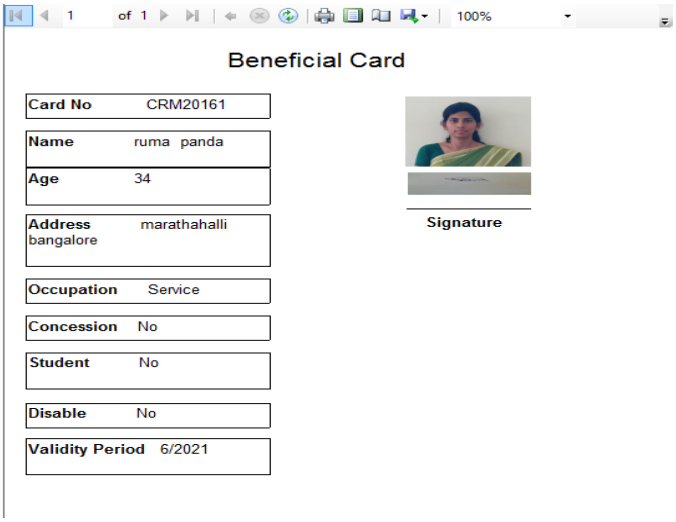

Figure 4: Beneficial Card

If user will give the print option then the beneficial card will be printed in the above form. The person will get the concession or not that will be specified in the card itself. Card validity will be specified in the card.

VI. Conclusions and Future Enhancements

Data mining techniques are used to analyze mass volume of data and detect hidden patterns and information from raw data. Here mainly Naive-Bayesian Classification has used to classify the customers for getting concession by generating the beneficial card. It will help the transportation company to analyze and forecast customer's pattern for generating the ID card to get concession. It can be applicable in case of health care system.

This application can be applied to other industries, retail industry, manufacturer industries and so on.

References

[1] J Han and M. Kamber, Data Mining concepts and Techniques (2nd ed) Morgan Kaufmann, 2006.
[2] Ngai,E.W.T.,Xiu,L.& Chau,D.C.K .”Application of data mining techniques in customer relationship management: A literature review and classification,” Expert system with application,2009,2,2592-2602
[3] Shu-hsien Liao, yin-ju chen, Hsin-hua Hsieh, ”Mining customer knowledge for direct selling and marketing ”Expert system with applications,2011,38,6059-6069
[4] Roung-Shiunn Wu,Po-Hsuan Chou,”Customer generation of multiple category data in ecommerce using a soft-clustering approach”, Electronic commerce Research and Applications, Volume 10,Issue 3,May- June 2011,Pages 331-441

[5] Kim,Y.H.&Moon,B.R.,"Multicampaign assignment problem ",IEEE Transactions on Knowledge and Data Engineering,2006,18,405-414
[6] Hyea Kyeong Kim,Jae Kyeong kim, Qiu Yi Chen "A Product network analysis for extending the market basket analysis", Expert Systems with Applications, Volume 39,Issue 8,15 June,2012,Pages 7403-7410
[7] S.U.Natchiar, S.Baulkani "Customer Relationship Management Classification using Data Mining Techniques " ,IEEE,2014.
[8] Abdullah S., Al-Mudimigh, Zahid Ullah, Farrukh Aleem Ddata Mining strategies And Techniques For CRM systems", System of Systems Engineering, 2009,SoSE 2009,IEEE International Conference.
[9] Ruey-Shun chen,Ruey-Chyi Wu and J.Y.Chen, "Data Mining Application in Customer Relationship Management of credit card Business", IEEE,2005
[10] S.Naveen, M.Venkataramana, "Customer relationship Management: Overview & Engineering Principles", IJSRP,Volume 4,Issue 5,may 2014.
[11] Gao Hua, "Customer Relationship Management Based on data Mining Technique", IEEE,2011.