Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 46 (2015) 725 – 731

Procedia Computer Science

International Conference on Information and Communication Technologies (ICICT 2014)

An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour

Femina Bahari T[a,]*, Sudheep Elayidom M[b]

[a]Research Scholar, Cochin University of Science & Technology, Kochi- 682022, India
[b]Associate Professor, Cochin University of Science & Technology, Kochi- 682022, India

Abstract

CRM-data mining framework establishes close customer relationships and manages relationship between organizations and customers in today's advanced world of businesses. Data mining has gained popularity in various CRM applications in recent years and classification model is an important data mining technique useful in the field. The model is used to predict the behaviour of customers to enhance the decision-making processes for retaining valued customers. An efficient CRM-data mining framework is proposed in this paper and two classification models, Naïve Bayes and Neural Networks are studied to show that the accuracy of Neural Network is comparatively better.

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)

Keywords: data mining framework; customer relationship management; prediction;classification.

1. Introduction

Data mining[15] is defined as a process that uses mathematical, statistical, artificial intelligence and machine-learning techniques to extract and identify useful information and subsequently gain knowledge from databases. Information technology tools, advanced internet technologies and explosion in customer data has improved the opportunities for marketing and has changed the way relationships between organisations and their customers are

* Corresponding author. Tel.: +91-9946012111;
E-mail address: feminabahari@gmail.com

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)
doi:10.1016/j.procs.2015.02.136

managed[3]. Customer Relationship Management helps in building long term and profitable relationships with valuable customers[7]. The set of processes and other useful systems in CRM help in developing a business strategy and this enterprise approach understands and influences the customer behaviour through meaningful communications so that customer acquisition, customer loyalty, customer retention and customer profitability are improved. The key factor in the development of a competitive CRM strategy is the understanding and analyzing of customer behaviour and this helps in acquiring and retaining potential customers so as to maximize customer value[4]. CRM-data mining framework helps organizations to identify valuable customers and predict their future. Each CRM element can be supported by various data mining models based on the tasks performed.

Data mining can be used in organizations for decision making and forecasting and one of the most common learning models in data mining that predicts the future customer behaviours[4] is classification. The prediction is done by the classification of database records into a number of predefined classes based on certain criteria[1, 8]. Neural networks, decision trees, naive bayes, logistic regression and SVM are the common tools used for classification.

To illustrate the performance of classification models we consider the CRM applications such as customer segmentation, prospecting and acquisition, affinity and cross sell, profitability, retention and attrition, risk analyses, etc[9] in banking domain. Instead of mass campaign banks focus on direct marketing campaigns as one measure to improve customer development. The banks use the data available to retain its best customers and to identify opportunities to sell them additional services[11]. Two classification models, the Multilayer Perception Neural Network (MLPNN) which have their roots in the artificial intelligence[5,13] and Naïve Bayes (NB) classifier, a simple probabilistic classifier based on applying Bayes' theorem [14] are used for the study.

This paper proposes an effective CRM-data mining framework and investigates the effectiveness of the two classification models in data mining in predicting the behaviour of customers in CRM application. An application in the bank direct marketing campaign is selected for the performance comparison of multilayer perception neural network and naïve bayes classifier. The data set is well known as bank marketing data from the University of California at Irvine (UCI)[11]. To assess the classifier performance, classification metrics such as accuracy rate, sensitivity analysis and specificity can be used[11]. The tool used for the study is Weka.

The remaining sections of the article are organized as Section 2 which covers the data used in the work and Section 3 specifies the Problem Statement. Section 4 describes the CRM-data mining framework, section 5 describes the main concepts used in the work, section 6 illustrates the implementation logic, section 7 analyzes the test results and finally section 8 concludes the article.

2. Data Sets Used

The dataset used for experiments in this paper, contains results of direct bank marketing campaigns[11]. It includes 17 campaigns of a Portuguese bank conducted between May 2008 and November 2010. The customer was offered a long-term deposit application by contact over phone. The dataset contains 45211 instances with two possible outcomes – either the client signed for the long-term deposit or not. In our experimental dataset 10% of the pre processed dataset is used and it contains 16 input variables. Eight variables relate to the client, four variables relate to the last contact of the current campaign and another four relate to the campaign:

- age, average yearly balance in Euros (numeric),
- job type, marital status, education (categorical),
- whether the client has credit in default, whether the client has a loan (binary),
- whether the client has a personal loan (binary).
- contact communication type, last contact month of year (categorical),
- last contact day of the month, last contact duration in seconds (numeric).
- number of contacts performed before and during this campaign and for this client (numeric),
- number of days that passed by after the client was last contacted from a previous campaign(numeric) ,
- outcome of the previous marketing campaign (categorical).

The output variable corresponds to campaign output, which has been reduced to a binary output which indicates whether the customer subscribes to a deposit scheme or not.

3. Problem Statement

To propose an efficient CRM-data mining framework for the prediction of customer behaviour in the domain of banking applications. Within the framework proposed, two classification models are studied and evaluated.

4. CRM-Data Mining Framework

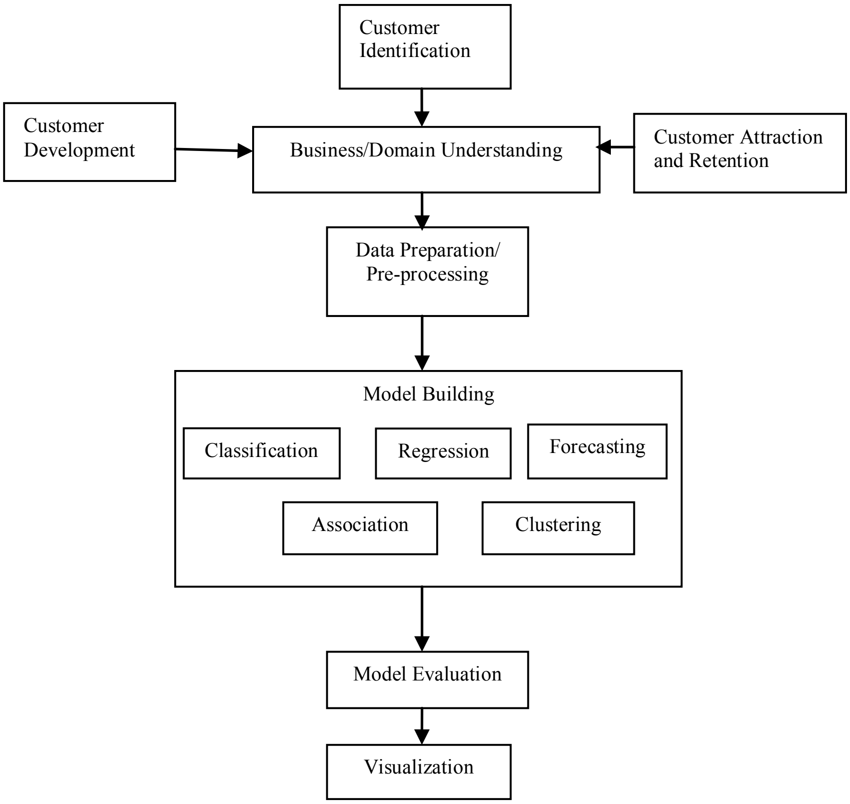

Fig. 1. CRM–data mining framework.

The proposed CRM-data mining framework is shown in Fig. 1. Understanding the business goals and requirements of the problem domain forms the initial phase of any problem in data mining. A close study and management of customer relationships and their interactions will help to identify attract and retain effective customers in the domain. The next phase of data preparation or preprocessing helps in preparing the data by the processes of cleaning, attribute selection, data transformation etc for further building up of models and their evaluation. Model construction in the CRM framework is a major step in which effective model to satisfy the business requirements is constructed. These models help in predicting the behaviour of the customers. Model evaluation and visualization measure the effectiveness of the model for enhancing their performance.

5. Concepts used

5.1. Multilayer Perception Neural Network (MLPNN)

Multilayer perception neural network (MLPNN) structure[5,13,16] is organized as a layered set of neurons. Among the input, output and hidden layers of neurons[5] the actual computations of the network are performed in the hidden layer, where each neuron sums its input attributes x_i after multiplying them by the strengths of the respective connection weights $w_{ij.}$ The activation function (AF) of this sum gives the output y_j and sigmoid function[16] is the AF used in the experiment.

$$y_j = f\,(\sum w_{ij,}\, x_i) \tag{1}$$

Back-propagation (BP) learning is the most common training technique used for MLPNN[2]. The sum of squared differences between the desired and asset value of the output neuron's E is defined as:

$$E = ½\sum_j (y_{d\,j} - y_j)\ 2 \tag{2}$$

Where y_j is the output of a neuron j whose desired value is $y_{d\,j}$. Weights $w_{i,j}$ in equation (1), are adjusted to finding the minimum error E of equation (2) as early as possible. The difference between the network outputs and the desired ones is reduced by the application of weight correction by BP. The neural networks helps in learning, and reducing the future errors[5]. Good learning ability, fast real-time operation, less memory demand, analysis of complex patterns are some of the advantages of MLPNN and the disadvantages include high-quality data requirement of the network, careful selection of variables a priori and so on[16].

5.2. Naive Bayes (NB)

Bayesian classifiers are helpful in predicting the probability that a sample belongs to a particular class. The technique is used for large databases because of its high accuracy and fastness to train with simple models. To estimate the parameters (means and variances of the variables) necessary for classification, the classifier requires only a small amount of training data. It also handles real and discrete data[5].

We can use Bayes' rule as the basis for designing learning algorithms, as follows: To learn some target function $f: U \rightarrow V$, or equivalently, $P\ (V|U)$, we use the training data to learn estimates of $P\ (U|V)$ and $P\ (V)$. Using these estimated probability distributions and Bayes' rule[14] new U examples can then be classified.

From a given set of training instances with class labels, a learner in classification learning problems, attempts to construct a classifier. The Naive Bayes classifier assumes all attributes describing U are conditionally independent given V. The number of parameters that must be estimated to learn the classifier is reduced dramatically by this assumption. For both discrete and continuous U[14], Naive Bayes is a widely used learning algorithm.

5.3. Weka

The Waikato Environment for Knowledge Analysis (Weka) is a machine learning toolkit used extensively for research, education and projects. Weka is introduced by Waikato University, New Zealand and is open source software written in Java (GNU Public License). It consists of collection of machine learning algorithms and tools for data mining tasks such as data pre-processing or data preparation, classification, association rules, clustering, regression, forecasting and visualization and is well suited for developing new machine learning schemes[16]. Weka 3.7.4 is used for experimentation in our work and can be run on Linux, Windows and Mac.

6. Implementation Logic

6.1. Data Preprocessing

Originally the dataset contained 79354 contacts and 58 attributes. Contacts with missing data or inconclusive results were discarded leading to 45211 instances and 17 attributes without missing values. Attributes relevant to contact information was obtained from the campaign reports whereas attributes relevant to clients were collected from the banks internal database. The output attribute is whether the client subscribes the bank term deposit or not. Out of 58 attributes there were several irrelevant attributes that adversely affects the data mining learning process. A manual feature selection procedure with the help of Rattle analysis was used to reduce the attributes to 29 input attributes and 1 output attribute. This was further reduced in various CRISP-DM iterations and finally 17 attributes including the output attribute were selected for the test data[11]. Only 10% of the instances (4521) available in the dataset were used for the experimental studies.

6.2. Modeling

All experiments were performed using weka tool and were conducted in windows 7 with Intel Core i3 2.53 GHz processor. In this work, we build two distinct DM classifier models: MLPNN and NB. For all the two models test mode of tenfold cross validation was used. MLPNN uses back propagation to classify the instances. The nodes are all sigmoid except for when the class is numeric. We set the number of hidden layers using the heuristic *a=round (M/2)* where *M* is the sum of attributes and classes. Other network parameters were set as follows: learning rate 0.3, momentum 0.2, training time 500ms and validation threshold 20. During the modeling phase we successfully tested the two models, MLPNN and NB using the weka tool. 10-fold cross validation was used to perform our experiments. Ten disjoint subsets were created by partitioning the original dataset randomly. Nine of the subsets were combined to form the training set and the remaining subset forms the testing set in each of the ten runs. Based on the response, two classes were obtained, those which responded positive and those responded negative.

7. Test Result

The results of our experiment for automatically classifying a given dataset are summarized in Table 1. which shows values for two different classifiers. For each method the classification accuracy (amount of correctly classified instances), true positive rate (the proportion of actual positives which are correctly identified as such), false positive rate (incorrectly classified positive), ROC area (area under the ROC curve) and the time taken to build the classifier model is shown.

Table 1. Results of comparison of classifiers, average over 10 runs.

Classifier	Classification Accuracy(%)	True Positive Rate(TPR)	False Positive Rate(FPR)	ROC Area	Time taken to build models (s)
MLPNN	88.63	0.41	0.052	0.847	1767.75
NB	87.97	0.47	0.067	0.858	0.08

MLPNN classifier model shows better accuracy (88.63%) among the two models experimented. NB gives better values of TPR (0.47), FPR (0.067) and ROC area (0.858). The time taken to build the model is very high for MLPNN (1767.75s).

In our work three statistical measures namely classification accuracy, sensitivity and specificity are used to evaluate the performance of the classification models. This information is given in terms of instances classified as true positive (TP), true negative (TN), false positive (FP) and false negative (FN)[6]. This information about actual and predicted classification defines a confusion matrix and is given in Table 2.

Table 2. Confusion Matrix for each classifier.

Classifier	Confusion Matrix		
	a	b	← Classified as
MLPNN	213	308	a = yes
	206	3794	b = no
NB	246	275	a = yes
	269	3731	b = no

Classification accuracy as shown in equation (3) is equal to the sum of TP and TN divided by the total number of cases N[5] and it refers to the ratio of the number of correctly classified cases.

$$Accuracy = \frac{TP+TN}{N} \tag{3}$$

Sensitivity in equation (4) is equal to ratio of TP to the sum of TP and FN and it refers to the rate of correctly classified positive (True Positive Rate).

$$Sensitivity = \frac{TP}{TP+FN} \tag{4}$$

Specificity in equation (5) is equal to TN divided by sum of TN and FP[6] and it refers to the rate of correctly classified negative (True Negative Rate).

$$Specificity = \frac{TN}{TN+FP} \tag{5}$$

MLPNN model shows the best values for accuracy (88.63%) and specificity (94.85%) for the training samples where as NB gives the best value for sensitivity (47.2%). The performance of the three classifiers is compared in terms of accuracy, sensitivity and specificity and the results of the comparison are as shown in Table 3.

Table 3. Performance comparison of classifiers in terms of accuracy, sensitivity and specificity

Classifier	Accuracy(%)	Sensitivity(%)	Specificity(%)
MLPNN	88.63	40.9	94.85
NB	87.97	47.2	93.28

Out of the 4521 instances in the dataset, the instances classified correctly and instances classified incorrectly for each model are as shown in Table 4. MLPNN classified 4007 instances correctly whereas NB classified 3977 instances correctly.

Table 4. Classification of 4521 instances in the dataset.

Classifier	Correctly classified instances	Incorrectly classified instances
MLPNN	4007	514
NB	3977	544

8. Conclusion and Future Scope

In this paper we propose an efficient CRM-data mining framework for the prediction of customer behaviour. Two classification models were used to predict the customer behaviour. In order to arrive at authentic research results it is always better to use standard bench marking datasets like UCI datasets. Hence we used the same in this work. The best model that achieves high predictive performance was MLPNN with accuracy rate of 88.63%. We also compared the performance of classifiers in terms of accuracy, sensitivity and specificity. This work can be extended to other new models like Neuro fuzzy classifiers, Ensemble of classifiers and so on. Also the same experimental set up can be applied to other huge live banking datasets.

References

1. Berson A, Smith S, Thearling K. Building data mining applications for CRM, McGraw-Hill; 2000.
2. Cortez P. Data Mining with Neural Networks and Support Vector Machines using the R/rminer Tool, *In Proceedings of the 10th Industrial Conference on Data Mining*, Germany: Springer; 2010. p. 572–583.
3. EWT Ngai. Customer relationship management research (1992–2002): An academic literature review and classification, *Marketing Intelligence, Planning;* 23, 2005. p. 582–605.
4. EWT Ngai, L Xiu, DCK.Chau. Application of Data Mining Techniques in Customer Relationship Management: A Literature Review on Classification, *Expert Systems with Applications;* 36- 2, 2009. p. 2592-2602.
5. Hany AE. Bank Direct Marketing Analysis of Data Mining Techniques, *International Journal of Computer Applications;* 85-7, 2014.
6. JW Han M Kamber. Data mining concepts and techniques, 2nd ed. Morgan Kaufmann, San Francisco, CA; 2006.
7. Ling, R., Yen D. Customer relationship management: An analysis framework and implementation strategies, *Journal of Computer Information Systems*; 41, 2001. p. 82–97.
8. Mi tra S, Pal SK, Mitra P. Data mining in soft computing framework: A survey, *IEEE Transactions on Neural Networks;* 13, 2002. p. 3–14.
9. MJA Berry, GS Linoff, Data Mining Techniques: For marketing, Sales and Customer Relationship Management, Indianapolis: Wiley; 2004.
10. S Moro, P Cortez, P Rita. A data-driven approach to predict the success of bank telemarketing, *Decision Support Systems;* 62, 2014. p. 23-31.
11. S Moro, R Laureano, P Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology, *Proceedings of the European Simulation and Modelling Conference*; Portugal, 2011 . p. 117-121.
12. Swift RS. Accelerating customer relationships: Using CRM and relationship technologies, N.J: Prentice Hall PTR; 2001
13. T Munkata. Fundamentals of new artificial intelligence, 2nd ed. London: Springer-Verlag; 2008.
14. Tom M Mitchell. Machine Learning, 2nd ed. McGraw Hill; 2010.
15. Turban E, Aronson JE, Liang TP, Sharda R. Decision support and business intelligence systems, 8th ed.Pearson Education; 2007.
16. Witten I, Frank E. Data Mining – Practical Machine Learning Tools and Techniques, 2nd ed. USA: Elsevier; 2005.