ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

Data mining: Techniques for Enhancing Customer Relationship Management in Banking and Retail Industries

P Salman Raju, Dr V Rama Bai, G Krishna Chaitanya

Research Associate, Dept. of I.T., Institute of Public Enterprise, Hyderabad, India

Professor, Dept. of I.T., MGIT, Dept. of I.T, Hyderabad, India

Assistant Professor, Dept. of I.T., HIST, Hyderabad, India

ABSTRACT: Currently several industries including like banking, finance, retail, insurance, publicity, database marketing, sales predict, etc are Data Mining tools for Customer Relationship Management. Leading banks are using Data Mining tools for customer segmentation and benefit, credit scoring and approval, predicting payment lapse, marketing, detecting illegal transactions, etc. The Banking and Retail industry is realizing that it is possible to gain competitive advantage deploy data mining. For retailers, data mining can be used to provide information on product sales direction, customer buying tradition and desires; etc. This article provides an critique of the concept of Data mining and Customer Relationship Management in organized Banking and Retail industries. It also discusses standard tasks involved in data mining; evaluate various data mining applications in different sectors.

Keywords: Data Mining, CRM, Analytical Intelligence, Banking and Retail Industries, clustering

I. INTRODUCTION

Data mining refers to computer-aided pattern discovery of previously unknown interrelationships and recurrences across seemingly unrelated attributes in order to predict actions, behaviours and outcomes. Data mining, in fact, helps to identify patterns and relationships in the data [1].

DM also refers as analytical intelligence and business intelligence. Because data mining is a relatively new concept, it has been defined in various ways by various authors in the recent past. Some widely used techniques in data mining include artificial neural networks, genetic algorithms, K-nearest neighbour method, decision trees, and data reduction. The data mining approach is complementary to other data analysis techniques such as statistics, on-line analytical processing (OLAP), spreadsheets, and basic data access. Data mining helps business analysts to generate hypotheses, but it does not validate the hypotheses.

II. RELATED WORK

1. Data Mining Defined Throughout Literature

1. Data mining is defined as the process of extracting previously unknown, valid, and actionable information from large databases and then using the information to make crucial business decisions – Cabena et al.
2. Data mining is described as the automated analysis of large amounts of data to find patterns and trends that may have otherwise gone undiscovered — Fabris.
3. The objective of data mining is to identify valid, novel, potentially useful, and understandable correlations and patterns in existing data — Chung and Grey

Objectives

1. Integrating retailer , suppliers and customers forgetter customer service
2. Descriptions of customer relations patterns
3. Constantly flexing the balance between marketing, sales and service inputs against changing customer needs to maximize profit Extracting or detecting hidden customer characteristics and behaviors from large databases
4. Development of online information kiosk for customers

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

2. Data mining techniques

Knowledge discovery in databases (KDD) is a CRM analytical tool which has received considerable attention in recent years (Frawley et al., 1992). This chapter provides a summary of the knowledge discovery process[3]. Moreover, since this research required the use of several data mining techniques, this chapter also includes a summary of the main data mining techniques used to assist analytical CRM.

Knowledge discovery: The KDD process is outlined in Figure 1. This process includes several stages, consisting of data selection, data treatment, data pre-processing, data mining and interpretation of the results. This process is interactive, since there are many decisions that must be taken by the decision-maker during the process. The stages of KDD process are briefly described below.

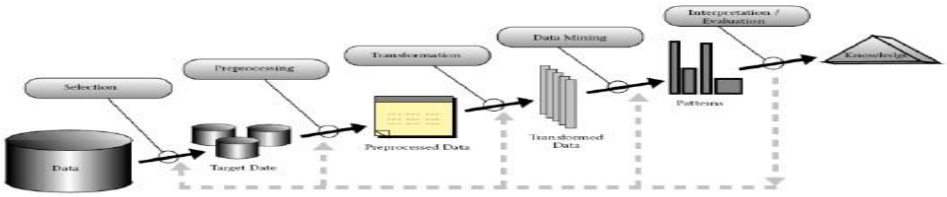

Figure1.KDD step by step process.

• **Data selection:** This stage includes the study of the application domain, and the selection of the data. The domain's study intends to contextualize the project in the company's operations, by understanding the business language and defining the goals of the project. In this stage, it is necessary to evaluate the minimum subset of data to be selected, the relevant attributes and the appropriate period of time to consider.

• **Data pre-processing:** This stage includes basic operations, such as: removing noise or outliers, collecting the necessary information to model or account for noise, deciding on strategies for handling missing data attributes, and accounting for time sequence information and known changes. This stage also includes issues regarding the database management system, such as data types, schema, and mapping of missing and unknown values.

. **Data transformation:** This stage consists of processing the data, in order to convert the data in the appropriate formats for applying data mining algorithms. The most common transformations are: data normalization, data aggregation and data discretization. To normalize the data, each value is subtracted the mean and divided by the standard deviation. Some algorithms only deal with quantitative or qualitative data. Therefore, it may be necessary to discredit the data, i.e. map qualitative data to quantitative data, or map quantitative data to qualitative data.

• **Data mining:** This stage consists of discovering patterns in a dataset previously prepared. Several algorithms are evaluated in order to identify the most appropriate for a specific task. The selected one is then applied to the pertinent data, in order to find indirect relationships or other interesting patterns.

• **Interpretation/Evaluation:** This stage consists of interpreting the discovered patterns and evaluating their utility and importance with respect to the application domain. In this stage it can be concluded that some relevant attributes were ignored in the analysis, thus suggesting the need to replicate the process with an updated set of attributes.

III PROPOSED WORK

3. Data mining in Customer Relationship Management

Customer relationship management (CRM) comprises a setoff processes and enabling systems supporting a business strategy to build long term, profitable relationships with specific customers. In figure2 Data mining can help companies

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

in better understanding of the vast volume of data collected by the CRM systems. In the past few years, many organizations (especially retailers and banks) have recognized the vital importance of the information they have on their customers.

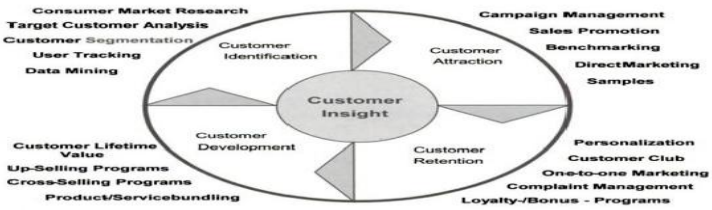

Figure2.CRM instruments

The banking industry is widely recognizing the importance of the information it has about its customers. Undoubtedly, it has among the richest and largest pool of customer information, covering customer demographics, transactional data, credit cards usage pattern, and so on. As banking is in the service industry, the task of maintaining a strong and effective CRM is a critical issue. To do this, banks need to invest their resources to better understand their existing and prospective customers. By using suitable data mining tools, banks can subsequently offer 'tailor-made' products and services to those customers.

3.1 Data mining and analytical CRM: Berry and Linoff (2000) defines data mining as the process of exploring and analyzing huge datasets, in order to find patterns and rules which can be important to solve a problem [17]. According to Ngai et al. (2009), association, classification, clustering, forecasting, regression, sequence discovery and visualization cover the main data mining techniques. In figure 3 these groups of data mining techniques can be summarized as follows:

•**Association** intends to determine relationships between attributes in databases (Mitra et al., 2002; Ahmed, 2004; Jiao et al., 2006). The focus is on deriving multi-attribute correlations, satisfying support and confidence thresholds [2]. Examples of association model outputs are association rules. For example, these rules can be used to describe which items are commonly purchased with other items in grocery stores.

• **Classification** aims to map a data item into one of several predefined categorical classes (Berson et al., 1999; Mitra et al., 2002; Chen et al., 2003; Ahmed, 2004). For example, a classification model can be used to identify loan applicants as low, medium, or high credit risks [4].

• **Clustering**, similarly to classification models, aims to map a data item into one of several categorical classes (or clusters). Unlike classification in which the classes are predefined, in clustering the classes are determined from the data. Clusters are defined by finding natural groups of data items, based on similitude marks or probability bulk models (Berry and Linoff, 2004; Mitra et al., 2002; Giraud-Carrier and Povel, 2003; Ahmed, 2004). For example, a clustering model can be used to group customers who usually buy the same group of products [5].

• **Forecasting** estimates the future value of a certain attribute, based on records' patterns. It deals with outcomes measured as continuous variables (Ahmed, 2004; Berry and Linoff, 2004). The central elements of forecasting analytics are the predictors, i.e. the attributes measured for each item in order to predict future behavior. Demand forecast is a typical example of a forecasting model whose predictors could be for example price and advertisement

• **Regression** maps a data item to a real-value prediction variable (Mitra et al., 2002; Giraud-Carrier and Povel, 2003). Curve fitting, modeling of causal relationships, prediction (including forecasting) and testing scientific hypotheses about relationships between variables are frequent applications of regression.

• **Sequence** discovery intends to identify relationships among items over time (Berson et al., 1999; Mitra et al., 2002; Giraud-Carrier and Povel, 2003). It can essentially be thought of as association discovery over a temporal database [7].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

For example, sequence analysis can be developed to determine, if customers had enrolled for plan A, then what is the next plan that customer is likely to take-up and in what time-frame.

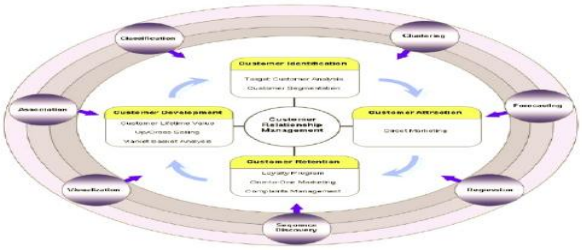

Figure3.Classification framework on data mining techniques in CRM

• **Visualization** is used to present the data such that users can notice complex patterns (Shaw, 2001). Usually it is used jointly with other data mining models to provide a clearer understanding of the discovered patterns or relationships (Turban et al., 2010). Examples of visualization applications include the mind maps [6].

IV RESULTS

4. Data Mining Applications in Banking Sector

Figure depicts the data mining techniques and algorithm that are applicable to the banking sector. Customer retention pays vital role in the banking sector. The supervised learning method Decision Tree implemented using CART algorithm is used for customer retention. Preventing fraud is better than detecting the fraudulent transaction after its occurrence. Hence for credit card approval process the data mining techniques Decision Tree, Support Vector Machine (SVM) and Logistic Regression are used. Clustering model implemented using EM algorithm can be used to detect fraud in banking sector.

4.1 Customer Retention in Banking Sector : Today, customers have so many opinions with regard to where they can choose to do their business Early data analysis techniques were oriented toward extracting quantitative and statistical data characteristics. To improve customer retention, three steps are needed: 1) measurement of customer retention; 2) identification of root causes of defection and related key service issues; and the 3) development of corrective action to improve retention.

1) ***Classification Methods:*** In this approach, risk levels are organized into two categories based on past default history. For example, customers with past default history can be classified into "risky" group, whereas the rest are placed as "safe" group.

Decision Tree: Decision trees are the most popular predictive models (Burez and Van den Poel, 2007). A decision tree is a tree-like graph representing the relationships between a set of variables [8]. Decision tree models are used to solve classification and prediction problems where instances are classified into one of two classes, typically positive and negative, or churner and non-churner in the churn classification case.

Figure 4. DM Techniques for Banking

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

Building a decision tree incorporates three key elements:
1- Identifying roles at the node for splitting data according to its value on one variable or feature.
2- Identifying a stopping rule for deciding when a sub-tree is created.
3- Identifying a class product for each absolute leaf node.

2) Value Prediction Methods*:* In this method, for example, alternative of classifying new loan applications, it attempts to predict prospect conventional amounts for new loan applications Neural Network and regression are used for this purpose. The most common data mining methods used for customer profiling are:

1. Clustering (descriptive)
2. Classification (predictive) and regression (predictive)
3. Association rule discovery (descriptive) and sequential pattern discovery (predictive)

4.2 Automatic Credit Approval using Classification Method: Fraud is a significant problem in banking sector. Detecting and preventing fraud is difficult, because fraudsters develop new schemes all the time, and the schemes grow more and more sophisticated to elude easy detection.

1) Classification Methods: Classification is perhaps the most familiar and most popular data mining technique. Estimation and prediction may be viewed as types of classification. There are more classification methods such as statistical based, distance based, decision tree based, neural network based, rule based [12].

C5.0: C5.0 builds decision trees from a set of training data in the similar way as ID3, using the concept of data entropy. The training data is a set S=S1,S2,.. of already classified samples. Each sample Si consists of a p-dimensional vector (x1,i,x2,i, …, xp,i), where the xj represent attributes or features of the sample, as well as the class in which si falls [13].

CART: A CART tree is a binary decision tree that is constructed by splitting a node into two child nodes repeatedly, beginning with the root node that contains the whole learning sample. Used by the CART (classification and regression tree) algorithm, Gini impurity is a measure of how often a randomly chosen element from the set would be incorrectly labeled if it were randomly labeled according to the distribution of labels in the subset.

Support Vector Machine (SVM: In machine learning, the polynomial kernel is a kernel function commonly used with support vector machines (SVMs) and other kernelized models, that represents the similarity of vectors (training samples) in a feature space over polynomials of the original variables. For degree-*d* polynomials, the polynomial kernel is defined as

$$K(x,y)= (x^T y+c)^d$$

Where x and y are vectors in the input space, i.e. vectors of features computed from training or test samples, c >0 is a constant trading off the influence of higher-order versus lower-order terms in the polynomial.

Logistic Regression:Logistic regression or logit regression is a type of regression analysis used for predicting the outcome of a categorical dependent variable based on one or more predictor variables. Instead of fitting the data to a straight line, logistic regression uses a logistic curve. The formula for a univariate logistic curve is

$$p = e^{c0+c1x1}/1 + e^{c0+c1x1}$$

To perform the logarithmic function can be applied to obtain the logistic function

$$log_e = \frac{p}{1-p} = c_0 + c_1 x_{10}$$

Logistic regression is simple, easy to implement, and provide good performance on a wide variety of problems [14].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

4.3 Fraud Detection in Banking Sector: Sometimes the given demographics and transaction history of the customers are likely to defraud the bank. Data mining technique helps to analyze such patterns and transactions that lead to fraud. Banking sector gives more effort for Fraud Detection. Fraud management is a knowledge-intensive activity. It is so important in fraud detection is that finding which ones of the transactions are not ones that the user would be doing.

1) The Clustering model: Clustering helps in grouping the data into similar clusters that helps in uncomplicated retrieval of data. Cluster analysis is a technique for breaking data down into related components in such a way that patterns and order becomes visible. This model is based on the use of the parameters' data cauterization regions.

In order to determine these regions of cauterization first its need to find the maximum difference ($DIFF_{max}$) between values of an attribute in the training data. This difference ($DIFF_{max}$) is split into $N_{interval}$ segments. $N_{interval}$ is the binary logarithm of the attribute values account N_{points}. In general, $N_{interval}$ can be found using another way of looking. Such calculation of $N_{interval}$ is based on the assumption that a twofold increase of N_{points} will be equal to $N_{interval}$ plus one.

Thus $N_{interval}$ centers and corresponding deviations that describe all values of the certain attribute from the training data appears.The final result of classification of the whole transaction is the linear combination of classification results for each parameter:

$$\text{Result} = w1 \times \text{Class1} + w2 \times \text{Class2} + \ldots + wn \times \text{Class n}$$

2) Probability density estimation method: To model the probability density function, Gaussian mixture model is used, which is a sum of weighted component densities of Gaussian form.

$$p(x) = \sum_{j=1}^{M} p(x \mid j)\, p(j)$$

The $p(x \mid j)$ is the jth component density of Gaussian form and the P(j) is its mixing proportion. The parameters of the Gaussian mixture model can be estimated using the EM algorithm (Computes maximum-likelihood estimates of parametersThe on-line version of the EM algorithm was first introduced by Nowlan.

$$P(j)\text{new} = \alpha\, P(j)^{\text{old}} + P(j \mid x)$$

Remembering that the new maximum likelihood estimate for P(j) is computed as the expected value of $P(j \mid x)$ over the whole data set with the current parameter fit.

4.4 Marketing: Bank analysts can also analyze the past trends, determine the present demand and forecast the customer behavior of various products and services in order to grab more business opportunities and anticipate behavior patterns. Data mining technique also helps to identify profitable customers from non-profitable ones. Another major area of development in banking is Cross selling i.e banks make an attractive offer to its customer by asking them to buy additional product or service.

4.5 Risk Management: Data mining technique helps to distinguish borrowers who repay loans promptly from those who don't. It also helps to predict when the borrower is at default, whether providing loan to a particular customer will result in bad loans etc. Bank executives by using Data mining technique can also analyze the behavior and reliability of the customers while selling credit cards too. It also helps to analyze whether the customer will make prompt or delay payment if the credit cards are sold to them.

V.DM IN RETAIL INDUSTRY

The retail industry is also realizing that it is possible to gain a competitive advantage utilizing data mining. For retailers, data mining can be used to provide information on product sales trends, customer buying habits and preferences, supplier lead times and delivery performance, seasonal variations, customer peak traffic periods, and similar predictive data for making proactive decisions.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

Marketing: One of the most widely used areas of data mining for the retail industry, as in the banking industry, is marketing. 'Market basket analysis' is a marketing method used by many retailers to determine optimal locations to promote products. Simply stated, it is the study of retail stock movement data recorded at a Point of Sale (PoS)—to support decisions on shelf space allocation, store layout, product location and promotion effectiveness. Knowing where to locate products and promote them effectively can increase store sales. Another marketing tactic employed by many retail stores is the use of 'loyalty' cards.

Risk Management: Risk management is another area where data mining is used in the retail industry. Previous purchasing patterns of customers are analyzed to identify those customers with low product or brand loyalty. Data mining enables retailers to remain competitive and reduce risks by helping them understand what their customers are really doing. Retailers can then target those customers who are more likely to buy a certain brand or product and also be able to promote products in stores where and when they are needed.

A majority of banks in developing countries (particularly in the public sector) are not usually known to exploit their information 'asset' for deriving business value through data mining and gain competitive advantage. But with progressive liberalization of rules on entry for private and foreign multinational banks, under the GATS framework of WTO, competitive pressure on domestic banks is increasing.

Fraud Detection: Retail industries must also be aware that fraud detection is absolutely necessary. It is estimated that 38% of retail shrink occurs because of dishonest employees. And with about 25 paise of every shrink Rupee traceable to PoS fraud, it is no wonder that retailers continue to look for ways to reduce the number of dishonest cashiers. Some supermarkets have begun to use digitized closed-circuit television (CCTV) systems, along with PoS data mining, to enable retail loss prevention managers to expose cashier stealing and sweet-hearting, assemble convincing evidence, and deal with these situations as a matter of routine. The managers decide what constitutes suspicious behavior and sends software to detect it. This is called 'exception-based reporting'.

Customer Acquisition and Retention: Data mining can also help in acquiring and retaining customers in the retail industry. The retail industry deals with high levels of competition, and can use data mining to better understand customers' needs. Retailer can study customers' past purchasing histories and know with what kinds of promotions and incentives to target customers.

V. CONCLUSION AND FUTURE WORK

Data mining is a tool used to extract important information from existing data and enable better decision-making throughout the banking and retail industries. They use data warehousing to combine various data from databases into an acceptable format so that the data can be mined. The data is then analyzed and the information that is captured is used throughout the organization to support decision-making. It is universally accepted that many industries (including banking, retail and telecom) are using data mining effectively. Undoubtedly, data mining has many uses in industries. Its practical applications in such areas as analyzing medical outcomes, detecting credit card fraud, predicting customer purchase behavior, predicting the personal interests of Web users, optimizing manufacturing processes etc. have been very successful. The retail industry is also realizing that data mining could give them a competitive advantage. Those banks and retailers that have realized the utility of data mining and are in the process of building a data mining environment for their decision-making process will reap immense benefit and derive considerable competitive advantage to withstand competition in future

REFERENCES

1. Frawley, W. J., Piatetsky-Shapiro, G., and Matheus, C. J. (1992). Knowledge discovery in databases: An overview. *AI Magazine*, 13(3):57.
2. Mitra, S., Pal, S., and Mitra, P. (2002). Data mining in soft computing framework: a survey. *IEEE Transactions on Neural Networks*, 13(1):3–14.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

3. Dr. Madan Lal Bhasin, 2006. Data Mining: A Competitive Tool in the Banking and Retail Industries
4. Berson, A., Smith, S., and Thearling, K. (1999). *Building Data Mining Applications for CRM*. McGraw-Hill, New York.
5. Ahmed, S. R. (2004). Applications of data mining in retail business. In *In- formation Technology: Coding and Computing, International onference on*, volume 2, page 455, Los Alamitos, CA, USA. IEEE Computer Society.
6. Shaw, M. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, 31(1):127–137.
7. Giraud-Carrier, C. and Povel, O. (2003). Characterising data mining software. *Intell. Data Anal.*, 7(3):181192.
8. Burez, J. and Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36:4626–4636.
9. K. Chitra, B.Subashini, Customer Retention in Banking Sector using Predictive Data Mining Technique, International Conference on Information Technology, Alzaytoonah University, Amman, Jordan, www.zuj.edu.jo/conferences/icit11/paperlist/Papers/
10. K. Chitra, B.Subashini, Automatic Credit Approval using Classification Method, International Journal of Scientific & Engineering Research (IJSER), Volume 4, Issue 7, July-2013 2027 ISSN 2229-5518.
11. K. Chitra, B.Subashini, Fraud Detection in the Banking Sector, Proceedings of National Level Seminar on Globalization and its Emerging Trends, December 2012.
12. K. Chitra, B.Subashini, An Efficient Algorithm for Detecting Credit Card Frauds, Proceedings of State Level Seminar on Emerging Trends in Banking Industry, March 2013.
13. Petra Hunziker, Andreas Maier, Alex Nippe, Markus Tresch, Douglas Weers, and Peter Zemp, Data Mining at amajor bank: Lessons from a large marketing application http://homepage.sunrise.ch/homepage/pzemp/info/pkdd98.pdf
14. Michal Meltzer, Using Data Mining on the road to be successful part III, http://www.dmreview.com/editorial/newsletter_article.cfm?nl=bireport& articleId =1011392&issue=20082, October 2004.
15. Fuchs, Gabriel and Zwahlen, Martin, What's so special about insurance anyway?, published in DM Review Magazine, http://www.dmreview.com/article_sub.cfm? articleId=7157, August 2003.
16. Dych, J. (2001). *The CRM Handbook: A Business Guide to Customer Relationship Management*. Addison-Wesley Professional, USA, 1 edition.