

# Customer Behaviour Analysis for Credit Card Proposers Based on Data Mining Techniques

R. ROSELIN

Assistant Professor of Computer Science,  
Sri Sarada College for Women (Autonomous), Salem – 16

C.HANUPRIYA

Assistant Professor of Computer Science,  
Sri Ganesh College of Arts and Science, Salem – 16

**Abstract**— Consumer financing have become increasingly important in the private sector for the last two decades. With the new reforms in the banking sector, the marketing of financial products has become very competitive, creating a need for strategizing the marketing efforts. This study investigates the shift of consumers towards the use of plastic money, with emphasis on credit cards. A survey of consumers holding one or no credit card was used for data collection. Variables related to demographics such as age, income level and gender have also been taken into consideration. Data mining techniques are useful to analyse the customer behaviour. Prediction is done using the relevant data set taken from the database on the basis of the attributes. Classification and Clustering algorithms are used to analyse customer behaviour towards credit card. Based on our observations, suggestions have also been made for managers to refine the target market.

**Keywords**—credit card, data mining, classification and prediction, survey, weka tool, customer behaviour, Customer relationship management

## I. INTRODUCTION

Customer Relationship Management is used to make more efficient business-customer relationships in order to maximize client satisfaction and thereby improve customer loyalty and retention. Data Mining grew from the persistent growth of techniques used to interrogate masses of data. Data mining creates the ideal environment for making customer relationship management (CRM). CRM is considered as the database marketing of an organization with the database of customers. CRM identifies the most profitable customers and provides the highest level of service to them. Customer-centric approach helps to improve customer satisfaction and thereby maximize customer retention. This paper mainly focuses on the behaviour of credit card proposers using classification and clustering algorithms implemented on real dataset collected through survey process.

Consumer behaviour means the study of individuals, groups or organizations about their process of selecting, securing, using and disposing the products, services, experiences or ideas to satisfy needs and the impact of this process on the consumer and the society. Behaviour concerns either with the individual or the group (e.g. In college friends influence what kind of clothes a person should wants to wears) or a firm (peoples working in firm make decision as to which products the firm should use.) The use of product is often so important to the marketer because this may influence how a product is best positioned or how we can encourage increased consumption. Consumer behaviour involves services and ideas as well as tangible products. The study of consumers helps firms and organizations to improve their marketing strategies by understanding issues such as how

1. The psychology of consumers that how he thinks, feels reasons and select between different alternatives.
2. The mindset of how the consumer is influenced by his or her environment.
3. The behaviour of consumer while shopping or making other marketing decisions.
4. How customer motivation and decision strategy differ between products that differ in their level of importance or interest that they entail for the customer; and
5. How management can adjust and improve their marketing campaigns and marketing ideas to more effectively reach customer [1].

Credit is a method of selling goods or services without the buyer having cash in hand. A credit card is only an automatic way of offering credit to a consumer. Today, every credit card carries an identifying number that speeds shopping transactions. To distinguish the two products debit card and credit card, credit card banks have offered their customers possibilities to borrow a limited amount with their card. This service or credit is also a way to attract them. However, even if customers have the possibility to revolve, not all of them use this service [2].

### A. Credit Card as A Transaction Medium

According to Zywicki, “over half and probably as much as 68% of credit card users should be considered “convenience users”, who use credit cards primarily as a transactional medium and who pay off their balances in full each month”. In addition, “60% of total bankcard volume generates no interest, up from roughly 50 percent six years ago” based on a study made by Visa in 2000 (Zywicki, 2000).

The two main reasons for using credit card as a transaction medium are:

- **To minimize their cash balances:** credit cards facilitate the possibility to shift their financial assets into investments more profitable.
- **Convenience:** credit card is a simple transaction medium, easy to use and widely distributed / accepted.

## B. Types of Credit Cards

- Credit cards now are of various types with different fees, interest rates and rewarding programs. When applying for a credit card, it is important to learn of their diverse types to know the one best suited to their lifestyle and financial status. Different types of credit cards available by banks and other companies/organizations are briefly described below.
  - **Standard Credit Card**
  - **Prepaid Credit Card**
  - **Secured Credit Card**
  - **Limited Purpose Credit Card**
  - **Charge Credit Card**
  - **Specialty Credit Card**
  - **Premium Credit Card**

In this paper the research has been carried over the credit card holders to find their behavior using data mining techniques such as classification and clustering algorithms. And comparative study has been made between these algorithms. To analyze the consumer behaviour real data set has been collected through survey process. In this data set 24 attributes were used to analyze their behaviour.

## II. LITERATURE REVIEW

### A. Review on Credit Card Based Data Mining Techniques

Rui Wu has discussed about promotion of credit card using data mining Based on Artificial Neural Network first the author introduces the concepts and applications related to the data mining domain. Then he applies the Back Propagation algorithm in Artificial Neural Network technology to a Credit Card promotion sample, and implements a program to analyze the sample [3].

According to the author Wei Wei, Data mining using neural networks can be applied in various financial fields such as risk mitigation, missing data filling, fraud detection, and customer profile classification etc. His paper work aims to develop methodologies to mine large sets of records and in particular to fill missing data in these records. The steps included data cleansing, data selection, data preprocessing, data representation, data clustering and finally the missing data filling. Furthermore, this work designs algorithms to evaluate the supervised neural networks performance, which is helpful for the future research on data prediction and classification. A case study based on a large data set of credit card records, which contains incomplete records, is performed to demonstrate that the proposed algorithms and their implementations accomplish the task of filling missing data in such records.

This work has the following contributions:

1. The methodology provides a mechanism to fill missing data in credit card record
2. Sets using neural network. Two-level filling procedures guarantee all "holes" filled with more accurate and reliable data.
3. Applications of the methodology finally get the complete clean, high quality data
4. Sets, which found a very solid foundation on credit card approval and fraud detection research.
5. The evaluation for neural network Back-Propagation [4].

### B. Review on Classification

Classification is one of data mining problems receiving great attention recently in the database community. Various classification algorithms have been designed to tackle the problem by researchers in different fields, such as mathematical programming, machine learning, neural networks and statistics.

K.V.Nagendra, and C. Rajendra, has analyzed the behaviour of the customer to improve the customer relationship with organization. They have used classification algorithm for their research. With the rampant competition in the domestic and international business, the Customer Relationship Management (CRM) has become one of matters of concern to the enterprise. CRM takes the customers as the center; it gives a new life to the enterprise organization system and optimizes the business process. In an effort to help enterprises understand their customers shopping behaviour and the ways to retain valued customers, data mining techniques has been proposed. As a rising subject, data mining is playing an increasingly important role in the decision support activity of every walk of life. This paper mainly focused on the research of the customer classification and prediction in commercial banks based on Naive Bayesian classifier that accommodates the uncertainty inherent in predicting customer behaviour. They have explained this with an example through a credit card dataset [5].

Masoumeh Zareapoor, Seeja.K.R, M.Afshar.Alam have discussed about the credit card fraud detection using various types of data mining techniques. The most commonly used fraud detection methods are Neural Network (NN), rule-induction techniques, fuzzy system, decision trees, Support Vector Machines (SVM), Artificial Immune System (AIS), genetic algorithms, K-Nearest Neighbor algorithms. This paper presents a survey of various techniques used in credit card fraud detection and evaluates each methodology based on certain design criteria.

They have explained about the properties of good fraud detection system as follows:

- It should identify the frauds accurately

- It should detecting the frauds quickly
- It should not classify a genuine transaction as fraud

In this paper, they have presented a comparative study of nine fraud detection methods based on credit card (Decision Tree, Neural Network, Bayesian Network, genetic algorithm, support vector machine, k nearest neighbor and Artificial Immune System, Hidden Markov Model, fuzzy neural network and fuzzy Darwinian system). The main objective of this paper is to review methodology of different detection methods based on credit card. They have considered the most important parameter in different methods such as, accuracy, speed and cost [6].

### III. DATA MINING TECHNIQUES

#### A. Data pre-processing methods

Raw data is highly susceptible to noise, missing values and inconsistency. The quality of data affects the data mining results. In order to help improve the quality of the data and consequently, of the mining results raw data is pre-processed so as to improve the efficiency and ease of the mining process. Data pre-processing is one of the most critical steps in a data mining process which deals with the preparation and transformation of the initial dataset. Data pre-processing methods are divided into following categories:

1. Data cleaning
2. Data Integration
3. Data Transformation
4. Data Reduction

#### B. Data Classification

Classification is the process of finding a set of models that describe and distinguish data classes and concepts, for the purpose of being able to use the model to predict the class whose label is unknown. Classification is a two step process, first, it build classification model using training data. Every object of the dataset must be pre-classified i.e. its class label must be known; second the model generated in the preceding step is tested by assigning class labels to data objects in a test data set. Each tuple/sample is assumed to belong to a predefined class, as determined by the class label attribute. The model is represented as classification rules, decision trees, or mathematical formulae.

Second step is model usage. It is for classifying future or unknown objects. It estimates accuracy of the model. The known label of test sample is compared with the classified result from the model. Model construction describe a set of predetermines classes. Accuracy rate is the percentage of test set samples that are correctly classified by the model. Test set is independent of training set, otherwise over-fitting will occur.

In this paper Classification algorithms such as Naïve baiyes, j48, jripper and ID3 are implemented in the weka tool and their accuracy has been evaluated.

#### C. Data Clustering

The process of making the congregation of abstract objects group as multiple clusters formed by similar objects is called Clustering. In the clustering process, one basic principle is maximizing the similarity in each cluster and minimizing the similarity between the various clusters. After clustering, the data objects in one cluster can be treated as a whole and have the common class label. Clustering is different from classification, the cluster's class attribute and the number of clusters are unknown before clustering on the data, or do not consider the data tuple with class label during study, instead that use clustering analysis to obtain the clustering class label based on the clustering result. Generally, the main clustering algorithms can be divided into the following categories:

##### 1. Partitioning method

This method first creates an initial division, then interactive through moving the object in the division interval to improve the partitioning. But this method can only find spherical clusters.

##### 2. Density-based method

If the density area only surrounds a threshold, it continues to cluster. This method can be used to filter "noise data" and to find arbitrary shape clusters.

##### 3. Grid-based methods

This method makes object be spaced into limits units. This method has a fast processing speed.

In this paper clustering algorithms such as simple k-means and Farthestfirst algorithms are implemented. A comparative study has been made among these algorithms.

### IV. EXPERIMENTAL RESULTS

A Comparative analysis is done for classification algorithms and clustering in terms of classification accuracy. The results for classification algorithms are reported in Table 5.1 and it is diagrammatically represented in Figure 6.1.

**Table 6.1 Comparative Analysis of Classification Algorithms**

S. No.	Classification Algorithm	Accuracy
1	Navie Bayes	89%
2	JRipper	94%
3	ID3	58%
4	J48	96%

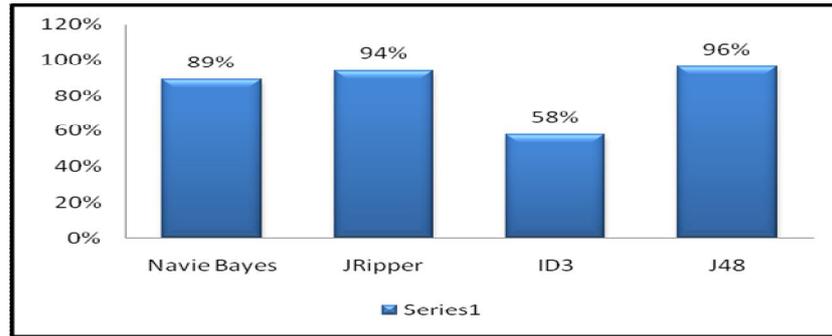


Figure 6.1 Bar chart for Classification Algorithms

The results of clustering algorithms are reported in Table 5.2 and it is diagrammatically represented in Figure 6.2.

Table 6.2 Comparative Analysis of Clustering Algorithms

S. No.	Clustering Algorithms	Accuracy
1	Simple K-Means	77%
2	Farthest First	78%

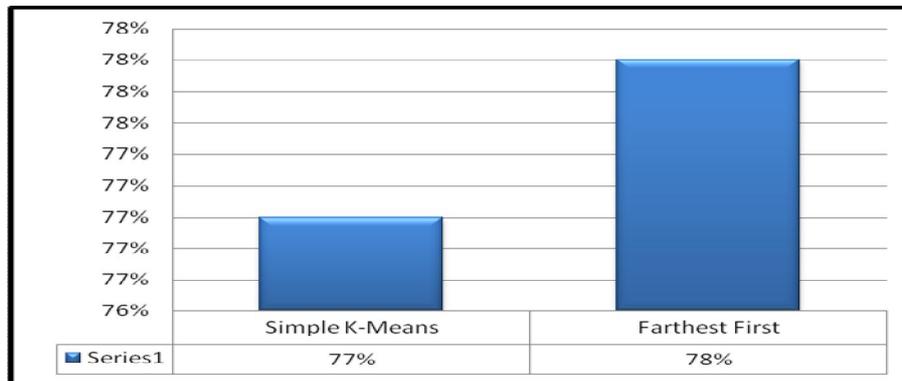


Figure 6.2 Bar chart for Clustering Algorithms

Classification algorithms and clustering algorithms are applied to the collected credit card dataset and a comparative study has been made to analyze the accuracy among the algorithms for the customer behavior. Among four classification algorithms j48 has given good accuracy of 96% comparing to other algorithms. When coming to the clustering algorithms Simple K-Means and Farthest First where compared with their accuracy and concluded that Farthest First gives 78% of accuracy where as Simple K-Means gives 77% of accuracy.

## V. CONCLUSIONS

Analysis of customer behavior enables companies to improve support of their customer oriented business processes, which aims to improve the overall performance of the enterprise. This paper focuses on getting more customer satisfaction. Data mining methodology has a tremendous contribution for researchers to extract the hidden knowledge and information. The research described in this paper also identified significant behavior for each segment. Data mining techniques can be used to motivate customers to increase their purchases and keep loyal to the company. The behavior of customers can be easily identified. Suppose commercial banks hope to increase the customers who will propose credit card with large number of valuable customer information in huge amounts of data accumulated by commercial banks, which is used to identify customers and provide decision support can apply the methods mentioned above such as classification and clustering. In this paper Customer behavior has been analyzed and comparative study is made between the classification and clustering algorithms. The classification algorithms such as Naïve Bayes, Jripper and J48 are applied for the credit card data set. J48 gives better accuracy than other two algorithms. Two clustering techniques have been used in this paper they are Simple K-means and FarthestFirst. It is concluded that security plays a vital role in proposing credit card. One who possess credit card for more than three years feels insecure in using credit card or plastic money. From this survey, it is understood that maximum number credit card proposers belong to urban region and their age is between 31 to 40. They use credit card for both foreign and domestic use.

REFERENCES

- [1] Abhijit Raorane & R.V.Kulkarni, *Data mining techniques: A source for consumer behavior analysis*.
- [2] Linda Delamaire., *Implementing a credit risk management system based on innovative scoring techniques*, the University of Birmingham ,February 2012.
- [3] Rui Wu.,” *Data Mining Based on Artificial Neural Network Analysis of Credit Card Promotion*”, University of California Santa Cruz, November 2012.
- [4] Wei Wei., *Data mining using neural networks for large credit card record sets*, Department of Computer and Information Science, May 2001.
- [5] Nagendra.K.V., and Rajendra.C., *Customer behaviour Analysis using CBA (Data Mining Approach)*, National Conference on Research Trends in Computer Science and Technology – 2012.
- [6] Afshar.Alam.M., Masoumeh Zareapoor., and Seeja.K.R., *Analysis of Credit Card Fraud Detection Techniques: based on Certain Design Criteria* , International Journal of Computer Applications, August 2012 .
- [7] Aman Paula., and Bhoj Raj Sharma., *Clustering Algorithms: Study and Performance Evaluation Using Weka Tool* , International Journal of Current Engineering and Technology, August 2013 .
- [8] Chien-I. Lee., and Shin-Mu Tseng and Kuo-Howang., *A Pre-Processing Method to Deal With Missing Values By Integrating Clustering And Regression Techniques*, 2003.
- [9] P.Isakki alias Devi., and Rajagopalan.S.P “*Analysis of Customer Behavior using Clustering and Association Rules* International Journal of Computer Applications, April 2012.
- [10] Pallavi., and Sunila Godara., *A Comparative Performance Analysis of Clustering Algorithms*, International Journal of Engineering Research and Applications (IJERA).
- [11] Dr.Raj Kumar., and Rajesh Verma., *Classification Algorithms for Data Mining: A Survey*, International Journal of Innovations in Engineering and Technology (IJET), August 2012.
- [12] Revathi.S., Dr..Nalini.T., *Performance Comparison of Various Clustering Algorithm*, International Journal of Advanced Research in Computer Science and Software Engineering , February 2013 .
- [13] Shrekar.S.S., Tina R. Patil., *Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification*, International Journal Of Computer Science And Applications, April 2013.