DATA MINING. CONCEPTS AND APPLICATIONS IN BANKING SECTOR

ADRIAN IONUT PASCU
PHD, UNIVERSITY OF CRAIOVA, DOCTORAL SCHOOL OF ECONOMICS SCIENCES,
e-mail:adrianionut.pascu

Abstract
The concept of banking refers to the multitude of services and products that commercial banks offer to clients and include besides transactional accounts both passive and active products. Due to the increased competitiveness in banking, the relationship between the bank and the client has become an essential factor for the strategy in order to increase customer satisfaction. Currently the banking system is able to store impressive amounts of data that they collect daily, from customer data and transaction details to data on their transactional or risk profile. The process through which large amounts of data are analyzed, extracted, identified and the information obtained using mathematical and statistical models are interpreted is known as data mining.

The discovery of knowledge from data involves identifying some models, some patterns with which certain events or possible risks are anticipated. This process helps banks to develop strategies in areas such as customer retention and loyalty, customer satisfaction, fraud detection and prevention, risk management, money laundering prevention.

The aim of this paper is to present the concept of data mining and the concept of data discovery (KDD), but also the impact and important use of data mining techniques in the banking sector. This paper explores and reviews various data mining techniques that are applied in the banking sector but also provides insight into how these techniques are used in different areas to make decision-making easier and more efficient.

Keywords: data mining, risk management, fraud prevention, CRM, KDD, marketing, money laundering.

Clasificare JEL: G21, G28

1. Introduction

In the past, all companies have focused on selling products and services without having or looking for knowledge about their own customers, but with increasing competition, attracting new customers has become more difficult and in this context, companies need to change their approach and step up their efforts to keep current customers. Social and economic evolution has changed consumer lifestyles, customers have become better informed and less inclined to respond to marketing communications. Thus, this change in consumers’ behavior has led companies to evolve, change their approach and have customer-oriented strategies, constantly improve service quality to ensure a good business relationship with customers. Companies have begun to invest in creating databases that can store an impressive amount of information for each customer, and adopt strategies to increase customer retention and development.

Banks have realized that customer relationships are very important for sustainable growth in the long run and Customer Relationship Management (CRM) is the strategy that can help build these long-term relationships and implicitly increase its revenue and profits. Developing the IT sector has helped the banking industry store stunning amounts of customer data, analyze and interpret this data in order to develop marketing strategies, risk, but also to improve processes and to better understand customers.

2. Data mining and discovery of data knowledge. Concepts

Decisions on marketing strategy, risk management and development are taken after "knowledge of large amounts of data is extracted" (Han and Kamber, 2011)[8], analyzed and
interpreted, this process being known as data mining. Berry and Linoff (Berry and Linoff, 2004)[2] define data mining as the process of exploring and analyzing large data sets to find patterns and rules that can solve a problem. The use of data mining is motivated by the need for decision makers' techniques in analyzing, understanding and viewing large amounts of data stored and collected from the data warehouse.

According to Berry and Linoff (Berry and Linoff, 2004)[2], the data mining process is often referred to as the Knowledge Discovery in KDD databases or knowledge creation. Fayyad et al. uses the term KDD that refers to the process of discovering useful knowledge from data, noting that data mining is a specific step in the data retrieval process (Fayyad, Piatetsky-Shapiro and Smyth, 1996b)[6]. Additional steps in the KDD process are data selection, data processing and transformation, data mining, and interpretation of the results, the whole process providing useful data creation (Figure 1).

![Figure 1. The process of discovering the knowledge of data](Fayyad, Piatetsky-Shapiro and Smyth, 1996a)[5].

Selecting data involves analyzing data and identifying the relevant subset of data to be selected for processing, data sources can be operational data or historical data. Data pre-processing involves the process of data cleaning and integration. Clearing means the removal of inconsistent or excessive data, and integration refers to the transformation of heterogeneous data extracted from multiple databases into homogeneous data. The transformation consists in processing and consolidating data in order to convert them into appropriate formats for the application of data mining algorithms. Data mining involves discovery of patterns in the dataset processed in the previous step, and more patterns can be found depending on applied techniques, all of these patterns need to be further analyzed to identify the most relevant. Interpretation is the stage in which discovered patterns are evaluated according to their utility, determining the degree of relevance in the applied field. You may also find that some values have been ignored in the analysis and then it is necessary to resume the process with an updated set of attributes.

According to the literature (Berry and Linoff, 2004)[2] among the most well-known data mining models are: classification; estimate; prediction; association rules; grouping or clustering; the description.
Classification is one of the most commonly used data mining techniques and consists of examining the features of a new object and assigning it a set of predefined classes. The purpose is to build on a model that can be applied to unclassified data in order to classify them. Example: Applying credit applications as high, medium, low risk depending on certain variables.

By estimating, unlike the classification that refers to definite results, the results obtained are continuously evaluated and used to assign a value to unknown variables. Example: A model is created by which each customer is assigned a value of 0 or 1 as the probability of buying a particular service or product. Then a threshold is set, and by ranking who is above this threshold is marked with 1, who is below that threshold is marked with 0.

Predictions are somewhat similar to estimation and classification, the difference between them is that the results are classified and depending on a future variable or future behavior, an outcome is estimated. Example: You can estimate the number of customers who will purchase a particular product or service as the service price changes.

Association rules are used to identify correlations between stored data, this technique being very effective in cross-selling. Example: Clients with an authorized account have a 60% probability of buying a credit card.

Grouping or clustering is the method of homogenous heterogeneous subgroup distribution. It is similar to the classification method, the difference being that if the groups are divided from the beginning into predefined classes in the group technique, they are divided as models and similar features are identified. Example: Clients are divided into groups as their trading habits are analyzed.

The description shows the existing data in order for users to see complex patterns, and is used with other data mining models to provide a better understanding of the discovered patterns.

3. Applications of data mining in banking

The data mining process analyzes data quantities, extracts and interprets the information obtained using mathematical and statistical models, thus helping companies benefit from a competitive advantage. Data mining techniques have a wide applicability in retail banking and are used in customer relationship and marketing management, risk management, fraud detection, money laundering prevention.

3.1. Marketing and Customer Relationship Management (CRM)

Marketing is the area where data mining techniques are most often used, and the banking sector is no exception. Client retention and customer portfolio growth in retail banking is increasingly difficult due to strong competition, and the main way to win or retain a customer is for a bank to be proactive, to know what customers want and to get up to expectations.

With data mining techniques, clients can be categorized according to needs, behaviors, value and other attributes. According to (Pulakkazhy and Balan, 2013)[15] there are two main models used for this classification, namely the credit scoring model and the behavioral scoring model. This classification allows banks to launch customer-oriented marketing campaigns tailored to each category and to offer different services and products for each category. For example, it is possible to determine how customers react to a new product or service launched, how they will react to a price increase on a particular product, or identify a customer category to increase cross-selling.

Data mining techniques applied to customer relationship management (CRM) systems can analyze customer data and find key indicators that help banks identify and anticipate their customers’ needs (Ngai, Xiu and Chau, 2009)[13], which allows them to have a targeted marketing strategy. The first CRM initiatives were launched in the early 1990s and were mainly focused on...
call center activities. The development of CRM and the enhancement of applicability has been strongly influenced by advances in information technologies, data management systems, improved analysis and communications, system integration, and internet development (Greenberg, 2001)[7].

CRM is of major importance for the banking system, being part of the strategy for building, managing and strengthening long-term relationships. The ultimate goal of CRM must be the personalized approach of customers and distinct entities by identifying and understanding their needs, preferences and differentiated behaviors. Nairn (Nairn, 2002)[12] defines CRM as a long-term business philosophy that focuses on collecting and understanding customer information, treating customers differently by providing a higher level of service to the best customers and using them together to increases customer loyalty and profitability. Although the benefits of using CRM are different in each business area, there are common benefits that are found in every sector of activity, including retail banking: lower customer acquisition costs, improved customer service, customer retention and loyalty, identifying profitable customers and increasing their profitability, increasing productivity.

Kracklauer et al. (Kracklauer, Mills and Seifert, 2004)[9] describes CRM as having 4 important customer relationship dimensions: customer identification, customer acquisition, customer retention, and customer development.

Client identification involves customer segmentation and target customer analysis, segmentation involves splitting a group of clients into smaller groups with similar features while client analysis involves defining the most attractive customer segments for a company based on their characteristics, and this selection process requires the collection of quantitative and qualitative data on customer groups.

Customer acquisition is the stage that follows identification and involves concentrating the company's efforts to attract the customers it wants. Using effective communication methods to these customers such as direct marketing, communicating competitive advantages such as price or other distinctive features of some products makes the acquisition of customers more efficient.

Customer retention is one of the main concerns of a company, satisfaction being the decisive factor that influences the degree of retention. We can define customer satisfaction by comparing their expectations with perceptions of service or product ownership, with the perceived value of the company. A positive interaction that the customer has with the company leads to a positive feeling and implicitly increases the client's loyalty to the service provider and products. Loyalty programs, bonuses, campaigns, or long-term relationships are examples of strategies used for customer retention.

Customer development is a priority for each company and aims to increase individual profitability by developing relationships and cross selling services and products.
Figure 2. CRM Components (Kracklauer, Mills and Seifert, 2004)[9].

Baran et al. (Baran, Galka and Strunk, 2008)[1] highlighted the importance of CRM and cross-selling and at the same time summarizes the main CRM objectives: identifying potential customers, understanding customer needs (current and dormant), differentiating profitable clients from unprofitable customers, reducing customer loss rates through increasing customer value and satisfaction, increasing the use of current products and services, increasing service quality and customer satisfaction, turning them from occasional customers into partners, integrating marketing and sales efforts across the various channels used by the company.

It is obvious that the use of data mining techniques in retail banking helps to develop more effective CRM strategies. By analyzing and interpreting data, different interactions take place with different clients that lead to personalized relationships with them and increased customer satisfaction.

3.2. Risk management

The most important area of application in which data mining techniques are used in retail banking is the risk assessment of lending to individuals and companies. Every credit decision taken by a commercial bank involves a certain degree of risk. Thus, granting a personal credit, a credit card, extending existing credit lines may be risky decisions for a commercial bank if it does not know its clients well. Data mining techniques are used to quantify the risk associated with credit facility decisions, making this process easier, limiting at the same time any financial losses to the bank.

According to Costa et al. (Costa, Folino, Locane, Manco and Ortale, 2007)[3], finding a pattern specific to a debt default behavior helps prevent future risks when similar patterns are discovered. Data mining techniques are applied in credit decision making and help estimate the probability of non-payment and assignment of credit score notes (Li and Liao, 2011)[11].

In the lending decision, banks determine for each client a behavioral score and a credit score. The behavioral score is associated with the probability of non-payment and is obtained on
the basis of past behavior by analyzing the history of the debt repayments entered in the credit bureau (BC), while the credit score is a value that is calculated taking into account several financial indicators and represents the customer’s creditworthiness.

Using data mining, it can be determined a credit behavior of borrowers with personal needs loans, mortgages, credit cards, using several variables including credit histories, duration of employment, residence, income amount, marital status. Customers with a history of credit behavior have a higher income, a stable employer is a good customer profile and are more likely to receive a loan to the detriment of customers who do not have a pay history debts and lower income. Historically stored data on customer debt payment can be used to develop a pattern of payment behavior for prospective customers requesting a credit. The determined score allows a bank to assess the client and decide whether the person is a good customer for a loan or if there is a risk of default.

3.3. Identifying and preventing fraud

Identifying fraud is a continuing and growing concern for the banking system, with data mining techniques being effective in detecting and preventing fraud. Fraud can be defined as an illegal mechanism that provides unauthorized funds or purchases goods and services without paying them.

When referring to the identification and prevention of fraud, we can relate to the splitting of two class transactions, legitimate or real transactions and illegal or fraudulent transactions (Ogwueleka, 2011)[14]. Detecting fraudulent transactions is a very complex process, and there is no system that predicts for sure whether a transaction is fraudulent or not, but just the probability of being an illicit transaction. In order for a system to be effective in detecting fraud, it must quickly and accurately identify frauds and not classify a real transaction as fraudulent (Zareapoor, Seeja and Alam, 2012)[17].

Cards are the main area where data mining fraud is highly effective, grouping method / cluster methods are used to classify transactions and surplus values, and they will be analyzed and used to prevent fraud (Dheepa and Dhanapal, 2009)[4]. Customer transaction data is recorded and analyzed, then the transaction patterns are established based on the results, and if significant deviations from these patterns are identified, alerts or warnings about the possibility of a fraudulent transaction are generated.

According to Rajdeepa (Rajdeepa and Nandhitha, 2015)[16], two approaches are used to detect fraudulent transactions. The former refers to the application of data mining techniques to data obtained from a third party (includes transactions from multiple companies) to identify fraudulent patterns, and then referring to these patterns with their own data to detect possible frauds. The second approach is based on identifying patterns exclusively in the bank's internal data.

Identifying and preventing fraud is a major and constant concern for the banking industry as it leads to financial losses, to the the risk of losing credibility to clients and to reputational risk.

3.4. Money laundering

Money laundering is the process by which it is hidden the illegitimate income originated and their transformation into apparently legitimate incomes. Money-laundering methods are becoming more and more sophisticated, and the activities that are being used to launder money more and more dangerous, from drug trafficking to terrorism financing, become a threat not only to the global financial system but also to the security of each country. The risks faced by the banking system are operational risks, reputational risks but also legal risk.

Taking all these aspects into account, governments and supervisory and regulatory authorities require financial institutions to implement systems, processes, procedures for combating
and detecting money laundering and terrorist financing. Traditional transaction analysis is not enough to discover patterns of more complex money laundering transactions, so data mining techniques are used to identify the suspicious transaction model that can detect money laundering transactions (Le Khac et al., 2009)[10]. In combating money laundering, the group method is used to divide and group transactions and cluster accounts from discovered similarities, this method helps build suspicious transaction patterns and detect customer or account risk models.

Failure to prevent and detect illegal transactions can have a major negative impact on banks, especially in terms of financial losses and reputational risk.

4. Conclusions

Data mining is the process of extracting knowledge from existing data and is used as a banking tool in order to find useful information in historical and operational data that can help in decision making.

Stages in this data mining process are data selection, integration and transformation, data mining, data evaluation and interpretation, and knowledge presentation.

Disclosed patterns help financial institutions anticipate events, identify potential risks, and massive investments in data mining make banks increasingly competitive.

The techniques used are effective in many areas of application such as customer relationship management, marketing, risk management, fraud detection and prevention, and money laundering detection.

5. Bibliography


