THE 6TH EDITION OF THE INTERNATIONAL CONFERENCE
E U R O P E A N   I N T E G R A T I O N
R E A L I T I E S   A N D   P E R S P E C T I V E S

Accounting and Financial Data Analysis. Data Mining Tools

Diana Elena Codreanu[1], Ionela Popa[2], Denisa Parpandel[3]

[1]*Constantin Brancoveanu University of Ramnicu Valcea, Faculty of Management Marketing in Economic Business, codreanudia@yahoo.com*

[2]*Constantin Brancoveanu University of Ramnicu Valcea, Faculty of Management Marketing in Economic Business, popaionela80@yahoo.com*

[3]*Constantin Brancoveanu University of Ramnicu Valcea, Faculty of Management Marketing in Economic Business, parpandeldenisa@yahoo.com*

Abstract: Computerized accounting systems in recent years have seen an increase in complexity due to the competitive economic environment but with the help of data analysis solutions such as OLAP and Data Mining can be a multidimensional data analysis, can detect the fraud and can discover knowledge hidden in data, ensuring such information is useful for decision making within the organization. In the literature there are many definitions for data mining but all boils down to same idea: the process takes place to extract new information from large data collections, information without the aid of data mining tools would be very difficult to obtain. Information obtained by data mining process has the advantage that only respond to the question of what happens but at the same time argue and show why certain things are happening. In this paper we wish to present advanced techniques for analysis and exploitation of data stored in a multidimensional database.

Keywords: knowledge discovery in database; multidimensional analysis; data mining tools

1. Introduction

In the current global economy, the performance of an organization is ensured and at the same time, conditional on the quality of decisions taken by its manager. Making the right choices based on a lot of data but also a complex process of analysis and synthesis thereof.

Necessary information to support decision may be obtained because of a powerful computer system, by which data is presented quickly, synthetic and relevant but also to provide opportunities for complex analysis and some predictions.

In analyzing the overall efficiency of an enterprise is very important information such as quantity and quality of existing accounts in the system at a time of economic enterprise, preferably at levels from becoming more analytical. (Crecană, 2000)

The decisions is the result of conscious choice of activities directions and engage in this action, which usually involves the allocation of resources. The decision resulting from the processing of information and knowledge, and belongs to a person or a group of people who have the necessary authority and responsible for the efficient use of resources in some given situation. (Filip, 2005)

In these circumstances, and computerized accounting systems have grown in complexity but with the help of data analysis solutions such as OLAP and Data Mining can be a multidimensional analysis of financial data - accounting, it can detect possible fraud and can discover knowledge hidden in data, trends can be established for certain indicators, thus ensuring useful information for decision making within the organization.

Manager, with financial and accounting information processed in a manner appropriate to its own needs, has the option to make an evaluation and Manage effectively the resources available to the company (we refer to financial, human and material) to meet its objectives and strategies set to medium term business development company.

For this, the financial manager has provided a number of tools such as:

- financial and accounting statements submitted and processed in accordance with internal reporting requirements, such as income statement, cash flow, financial statement;
- analysis of the income, expenses and profits;
- determining the minimum threshold of profitability;
- analysis of receipts and payments.

EF Codd, considered the father of relational databases, said in 1993 that "the company's ability to compete successfully and thrive is directly correlated with the effectiveness of its OLAP capability."

2. Data Mining

The Data Mining means the extraction process which takes place in the knowledge database based on some information requirements and to validate the information obtained. This is the approach that has been accepted more and more lately.

Date mining became known in the '90s, when speaking of data mining or mining the data "in many environments, whether it is academic whether the business.

In 1997 Pregibon - Research Scientist Google Inc.., says that "Data mining is a mixture of Statistics, IA (Intelligence Artificial) and database research" (Pregibon, Data Mining, Statistical Computing and Graphics Newsletter, 7, p.8, 1997).

The literature abounds with definitions of the concept of data mining. Han and Kamber, in his book "Data Mining - Concepts and Technique" from 2006 suggest the following definition "a lot of techniques that can be used to extract valuable information and knowledge from massive volumes of data."

Gray and Watson believe that "data mining allows analysts and store managers to find the answers to company data, which they have not even asked[1]."

The Knowledge Discovery System that can work on a large database system is called Knowledge Discovery in Databases System - KDD.

Between KDD (Knowledge Discovery in Databases) and Data Mining there are authors who differentiation (such as Fayyad). (Fayyad, et ali., 1996) Data mining systems are seen as a query technology, which involves the discovery and extraction of patterns, trends in the data using algorithms. Systems of knowledge discovery in databases it involves the entire process of discovering

---

[1] Gray P. H.J. Watson (1996) *The new DSS OLAP, MDD and: Data Warehouses, KDD.*

useful knowledge from data and besides we can say that Data Mining includes: problem definition, data collection and data interpretation.

It often happens that data mining is associated with: SQL queries, data retrieval, analysis of multidimensional database systems using OLAP tools, reports and graphs used for presenting data, traditional statistical data processing. This is because the combination of data mining, many times it happens to be used with traditional techniques of query or data analysis.

Although the above techniques are used together, yet they do not realize the same thing. Data analysis techniques and views have as main objective the verification of hypotheses, while data mining is to obtain answers to questions like *"What are the main causes that generate this phenomenon?"* or "*How I get some results?* ".
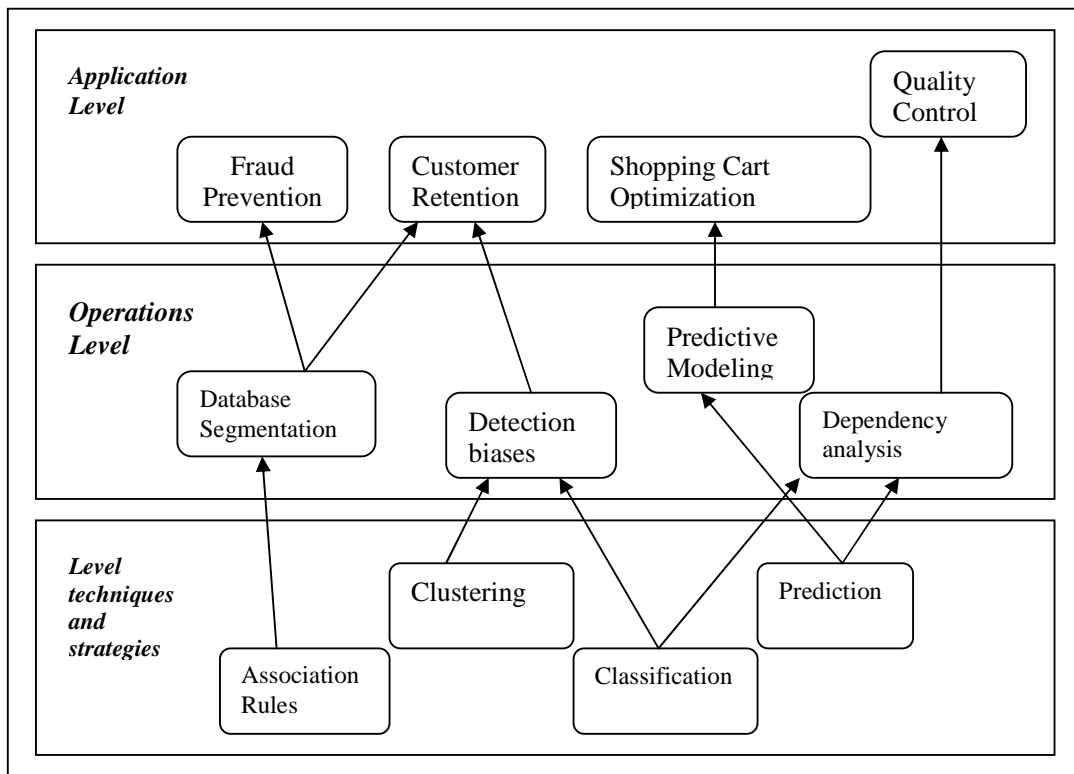

Figure 1: Applications, operations, techniques and strategies

At the organizational level can identify three levels where data mining is done as follows:

- *Application Level*: involves the use of one or more strategies and are used in decision making level;

- *The operations*: at this level using one or more technical level, and use information;

- *The techniques and strategies:* involves the use of specific data mining tools to extract knowledge from data.

3. The Process of Data Mining

The process of data mining is a process that involves going through the following successive steps:

1. Data collection and preparation process follow undergoing analysis. Data to be used in data mining process may have several sources: text files, spreadsheets, file type (Excel), relational databases, data warehouses. After collection, they are subject to a cleaning process to be used in the analysis. *Data analysis or application of an algorithm or a method of data mining.* In order to analyze the data, choice of data mining methods should be taken into account: the type of learning (supervised or unsupervised) that existing cases be used to construct the model and will be used to model testing Which attributes will be used for analysis.

3. The interpretation algorithm results: output is to examine the data mining tools to assess whether the knowledge obtained is useful or not. If the data obtained satisfactory then the process of data mining be resumed based on new attributes and comments.

4. Using the results of the analysis process to address new problems.

4. Strategies of Data Mining

Using a strategy of data mining approach defines a solution to solve a problem. Any technique used for data mining uses a strategy. Data mining strategies are divided according to year two major categories of learning methods:

- supervised strategies;
- unsupervised strategies.

With supervised learning models are made based on input attributes to predict the output attributes. Output attributes as the wear and dependent variables because their value is influenced byamount one or more input attributes. In turn, the input attributes is called the independent variables. In practice, we encounter supervised learning algorithm using one or more attributes of the output (dependent variable).

Where we unsupervised learning in the model type attributes are not found output, so all attributes will be of type input, independent variables.
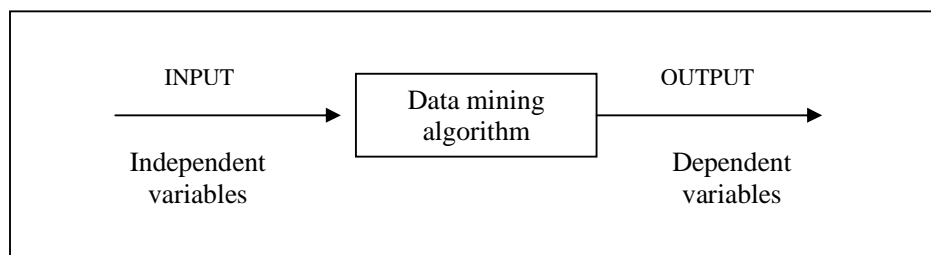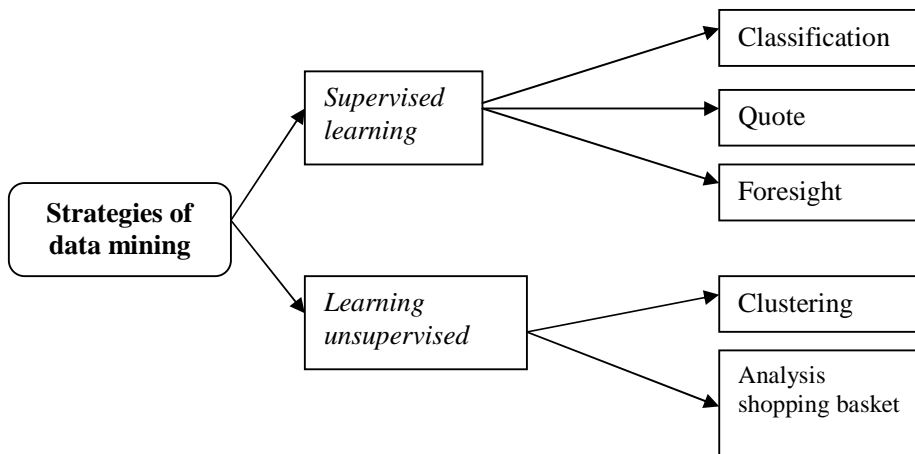
Figure 2: Supervised data mining algorithm

(Sandu, 2009)

Figure 3: Strategies for Data Mining

4.1. Strategies Supervised Data Mining

1. Classification

In the literature this strategy is known as the categorization. Classification is regarded as the most common and easiest to understand. Classification process is based on four fundamental components (Gorunescu, 2006):

▪ *Class* (dependent variable model) is a categorical variable. Sample classes: myocardial infarction, class of stars (galaxies), the class of an earthquake (hurricane), etc.

▪ *The predictors* (independent variables of the model) are given the characteristics of the data subject classification process. Examples of predictors: smoking, blood pressure, satellite images, geological records specific season, Loctite produce the phenomenon, etc.

▪ *The set of training* (learning) - Training Data Set: dataset is given by values containing the first two components, the model being used to identify the appropriate grade on the grounds of the predictors. Examples include: the group of patients tested for heart attack, earthquake research database containing images and telescopic tracking.

▪ Testing data set – in composition to enter new data will be subject classification model previously designed and thus can determine the model performance (classification accuracy).

Models were constructed by the algorithm can then be used to classify new cases.

Several types of situations you can use the classification problem:

- Profiling a person considered "successful";

- Establish carateristicilor under which people who have suffered a myocardial infarction to be distinguished from those who have not suffered a heart attack;

- Develop a profile that can help us to differentiate between men and women suffering from heart disease;

- Establish geographical areas at risk for an earthquake.

Classification, and strategy data mining, is used to analyze the current behavior and not to the future. For example, if realized a financial model for buying a car, then this strategy may help determine whether the buyer has low or high credit risk now and not in the future. In order to determine future behavior predictive models are used.

The category most commonly used methods for classification can include:

- classification trees / decision;
- statistical analysis;
- neural networks;
- Bayesian classifiers;
- genetic algorithms;
- classifiers based on rules

2. Quote

The main objective of the estimate, as for the classification is to determine the output value of an unknown attribute (dependent variable). Like the difference between the two strategies can estimate states that the output is numerical attributes, while the classification, as described above, output is categorical attributes.

Examples of situations where it can be used for estimation:

- estimate the probability that a person suffering a heart attack;
- estimate of income a person owning a second car;
- estimate the probability that a bank card and stealing.

A significant category of data mining techniques are used to solve problems of classification and estimation, but are not able to solve both simultaneously. Where existing data mining tool supports only one strategy over the other, the situation can be resolved so that the problem can be solved using existing techniques. It appeals to transform categorical attributes numeric attributes, or vice versa by converting numerical attributes categorical attributes.

3. Forecast

The first two strategies presented, classification and estimation, are currently used to predict behavior while the third strategy, the forecast is used to estimate future behavior. This is the major difference between the three supervised learning strategies. In the case of a forecast model output attributes can be categorical or numerical.

For example, prediction strategy can be used to determine whether telephone subscribers from a firm plan to change service provider in the next period.

In practice, many data mining techniques are used appropriately for classification and estimation, may be suitable for estimation. Choice model to be used correctly for classification, estimation or prediction is given by nature to be subject to data analysis process.

4.2. Unsupervised data mining strategies

1. Non-supervised clustering

Clustering is the method of dividing a dataset into several parts (clusters). If non-supervised clustering dependent variable does not meet a target to help learning. Knowledge structure learning program is designed by reference to the cluster quality metrics for the division comments in one or more clusters. At the beginning of learning the number of clusters is known as clustering to identify conceptual structures in data.

Clustering is used when you want:

- determining outlier data (data of abnormal cases);
- evaluating the results of a supervised learning model
- establish meaningful set of attributes used for supervised learning

Clustering technique helps determine where abnormal data, cases are known as outlier. Non-statistical techniques of data mining to identify abnormal cases while the applications used for data mining statistical data exclude these abnormalities in the data set. For example, identifying outlier data are used to determine whether a credit card is fraudulently used or not.

2. Analysis shopping basket

With shopping basket analysis can discover non-intuitive relationship between the products are sold. Following this analysis, the results obtained, the trader can make better decisions in terms of promotional campaigns, presentation of products in various catalogs and offers strategy by which to present their products on the shelves.

5. Conclusions

Using data mining analysis to obtain new knowledge without the need for human intervention. Practical data mining analysis are focused on the discovery of new knowledge from data.

Information obtained by using data mining techniques to be valid. The accuracy and completeness are the two characteristics of the underlying validity of the data. This information should not only be valid and the data mining process itself.

Another feature of the information obtained by using data mining techniques is operationality, is based on information obtained, the organization can start various actions to ensure a number of advantages. It happens that in many cases, the results of data mining techniques are not very easy to apply. With data mining techniques to historical data and then process it may happen that the results may not be current.

6. Acknowledgement

6. Bibliography

Crecană, C. (2000). *Rentabilitatea întreprinderilor mici şi mijlocii/Return to small and medium* Bucharest: Economic.

Fayyad, U.M.; Piatetsky-Shapiro, G.; Smyth, P., &Uthurusamy R. (1996). *Advances in Knolwledge discovery and Data Mining.* AAAI Press/Mit Press, menlo Park, Californian USA

Filip F. G. (2005). *Decizie asistată de calculator: decizii, decidenţi, metode de bază şi instrumente informatice /Computer Aided Decision: Decision makers, basic methods and associated tools*. Bucharest: Technical Publishing House.

Gray P. & Watson, H.J (1996). *The new DSS: Data Warehouses*, OLAP, MDD and KDD.

Gorunescu, Florin (2006). *Data Mining. Concepte, modele şi tehnici/Data Mining. Concepts, models and techniques*. Cluj-Napoca: Blue House.

Sandu, Daniela Ioana (2009). *Soluţii informatice pentru inteligenţa afacerilor / Solutions for business intelligence*. Bucharest: Thesis.