

Data Mining Solutions for the Business Environment

Ruxandra PETRE

University of Economic Studies, Bucharest, Romania

ruxandra_stefania.petre@yahoo.com

Over the past years, data mining became a matter of considerable importance due to the large amounts of data available in the applications belonging to various domains. Data mining, a dynamic and fast-expanding field, that applies advanced data analysis techniques, from statistics, machine learning, database systems or artificial intelligence, in order to discover relevant patterns, trends and relations contained within the data, information impossible to observe using other techniques.

The paper focuses on presenting the applications of data mining in the business environment. It contains a general overview of data mining, providing a definition of the concept, enumerating six primary data mining techniques and mentioning the main fields for which data mining can be applied. The paper also presents the main business areas which can benefit from the use of data mining tools, along with their use cases: retail, banking and insurance. Also the main commercially available data mining tools and their key features are presented within the paper.

Besides the analysis of data mining and the business areas that can successfully apply it, the paper presents the main features of a data mining solution that can be applied for the business environment and the architecture, with its main components, for the solution, that would help improve customer experiences and decision-making.

Keywords: *Data mining, Business, Architecture, Data warehouse*

1 Introduction

Nowadays, companies collect huge volumes of data on a daily basis. Analyzing this data and discovering the meaningful information contained by it became an essential need for businesses.

As the business environment develops and changes constantly, facing every day new challenges, the companies try to strengthen their market position and achieve competitive advantage by using new and innovative solutions, like data mining.

Data mining solutions implement advanced data analysis techniques used by companies for discovering unexpected patterns extracted from vast amounts of data, patterns that offer relevant knowledge for predicting future outcomes.

2. General overview of data mining

The availability and affluence of data belonging to various domains make data

analysis a matter of significant importance and necessity today. Data mining – the analysis step within the KDD (Knowledge Discovery in Databases) process – uses a diversity of advanced data analysis methods to explore the data and discover useful patterns and trends.

Data mining consists of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns (or models) over the data. [1]

With the imminent growth of the amounts of data in every application, using data mining methods for automatically identifying valid and meaningful patterns in order to produce useful information and knowledge became a requirement for various fields including business, education or science and engineering, fields for which data mining can fulfill the following purposes:

- *Business* – data mining can be applied in retail, banking or insurances, for activities like customer segmentation

and retention, market basket analysis or fraud detection;

- *Education* – data mining can be applied for grouping students, predicting student performance, planning and scheduling courses or understanding student behavior;
- *Science and engineering* – data mining can be used for domains like bioinformatics, astronomy, medicine, genetics, electrical power, telecommunications or climate data.

Data mining can be defined as a process of exploring and analysis for large amounts of data with a specific target on discovering significantly important patterns and rules. Data mining helps finding knowledge from raw, unprocessed data. Using data mining techniques allows extracting knowledge

from the data mart, data warehouse and, in particular cases, even from operational databases. [2]

The data mining methods, used for extracting hidden patterns from the data, are classified into the following two categories: description methods and prediction methods. Description methods are oriented to data interpretation, which focuses on understanding (by visualization for example) the way the underlying data relates to its parts. Prediction-oriented methods aim to automatically build a behavioral model, which obtains new and unseen samples and is able to predict values of one or more variables related to the sample. [3]

Data mining analyzes the data by applying a wide variety of techniques, developed for the efficient handling of large volumes of data. The six primary data mining techniques are presented below in figure 1:

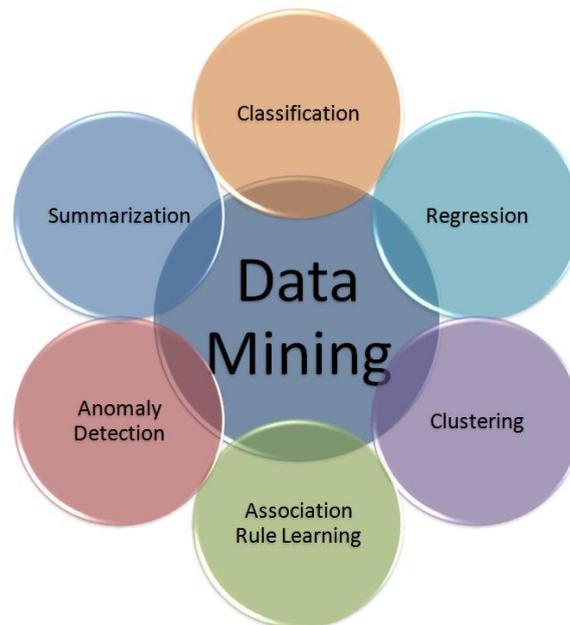


Fig. 1 Data mining techniques

The main data mining techniques are organized into the following categories: [1]

- *Classification*: consists of a function that maps (classifies) a data item into one of several predefined classes;
- *Regression*: involves a function that maps a data item to a real-valued

prediction variable;

- *Clustering*: is a common descriptive task where one seeks to identify a finite set of categories or clusters to describe the data;
- *Association rule learning (Dependency modeling)*: consists of finding a model that describes significant dependencies between variables;

- *Anomaly detection (Change and deviation detection)*: focuses on discovering the most significant changes in the data from previously measured or normative values;
- *Summarization*: involves methods for finding a compact description for a subset of data.

Data mining has evolved in the past two decades, becoming a fundamental discovery process. It has incorporated techniques from many other fields, including statistics, machine learning and database systems.

The diversity of data and the multitude of data mining techniques provide various

applications for data mining, which have improved many domains of human life.

3. Data mining applications for business

Data mining is defined as a business process for exploring large amounts of data to discover meaningful patterns and rules. [4] Companies can apply data mining in order to improve their business and gain advantages over the competitors.

The most important business areas that successfully apply data mining, presented in Fig. 2 below, are:

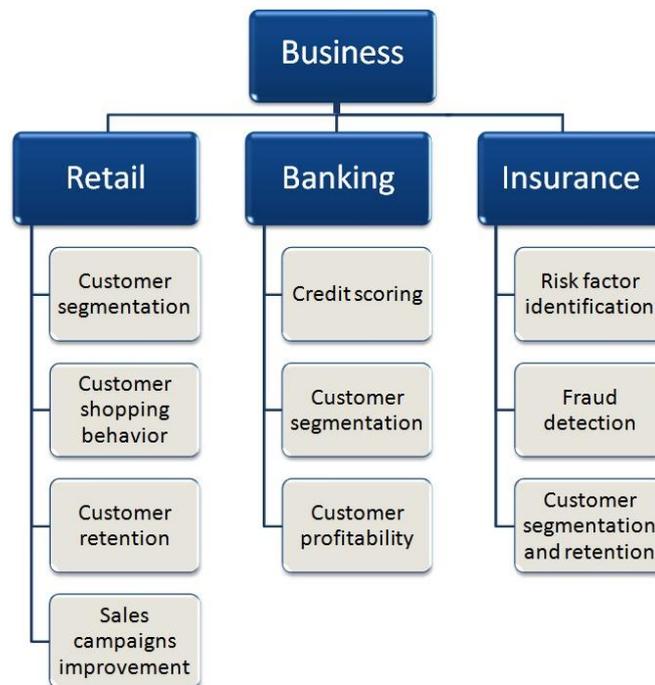


Fig. 2 Business areas that successfully apply data mining

1. Retail

Retail data mining can help identify customer buying behaviors, discover customer shopping patterns and trends, improve the quality of customer service, achieve better customer retention and satisfaction, enhance goods consumption ratios, design more effective goods transportation and distribution policies, and reduce the cost of business. [5]

Data mining techniques have many applications in the retail industry, including the following:

- *Customer segmentation*: identify customer groups and associate each customer to the proper group;
- *Establish customer shopping behavior*: identify customer buying patterns and determine what products the customer is likely to buy next;
- *Customer retention*: identify customer shopping patterns and adjust the

product portfolio, the pricing and the promotions offered;

- *Analyze sales campaigns*: predict the effectiveness of a sales campaign based on the certain factors, like the discounts offered or the advertisements used.

Retail industry offers a wide area of applications for data mining due to the large amounts of data available for companies.

2. Banking

There are various areas in which data mining can be used in financial sectors like customer segmentation and profitability, credit analysis, predicting payment default, marketing, fraudulent transactions, ranking investments, optimizing stock portfolios, cash management and forecasting operations, high risk loan applicants, most profitable Credit Card Customers and Cross Selling. [6]

The main examples of applications of the data mining techniques in the banking industry are the following:

- *Credit scoring*: distinguish the factors, like customer payment history, that can have a higher or lower influence over loan payment;
- *Customer segmentation*: establish customer groups and include each new customer in the right group;
- *Customer retention*: identify customer shopping patterns and adjust the product portfolio, the pricing and the promotions offered;
- *Predict customer profitability*: identify patterns based on various factors, like products used by a customer, in order to predict the profitability of the customer.

The information systems for the banking industry contain large amounts of operational and historical data, being a fitted application area for data mining.

3. Insurance.

Data mining can help insurance firms in business practices such as: acquiring new customers, retaining existing customers, performing sophisticated classification or correlation between policy designing and policy selection. [7]

In insurance the data mining techniques have the following applications:

- *Risk factor identification*: analyze the factors, like customer claims history or behavior patterns, that can have a stronger or weaker influence over the insured's level of risk;
- *Fraud detection*: establish patterns of fraud and analyze the factors that indicate a high probability of fraud for a claim;
- *Customer segmentation and retention*: establish customer groups and include each new customer to the appropriate group and identify discounts and packages that would increase customer loyalty.

Data mining techniques have many applications in the insurance business and can improve it by analyzing the large amounts of data available for companies.

4. Data mining tools used in the business environment

Data mining tools commercially available implement various data mining techniques for performing advanced data analysis on large volumes of data. The main data mining products, presented in Table 1 below, along with their key features, are: IBM SPSS Modeler, developed by IBM, the data mining tools included by Microsoft SQL Server Analysis Services, Oracle Data Mining, embedded within the Oracle database, SAS Enterprise Miner, produced by SAS, and STATISTICA Data Miner, developed by StatSoft.

Table 1 Main commercially available data mining tools and their key features

Data mining tool	Key features
IBM SPSS Modeler	<ul style="list-style-type: none"> • Data mining and text analytics software application used for building predictive models • Intuitive graphic user interface that allows users to import, manage and analyze their data • Data mining techniques included are: clustering (K-means, Support Vector Machine), classification (Bayesian networks, regression, neural networks, decision trees), association rules (Apriori), anomaly detection • Application for which it can be used: forecasting sales, customer relationship management, risk management or fraud detection
Microsoft SQL Server Analysis Services	<ul style="list-style-type: none"> • OLAP, data mining and reporting tool in Microsoft SQL Server • Used to create, manage, and explore data mining models, and then create predictions by using those models • Data mining algorithm types included are: classification, regression, clustering, association algorithms, sequence analysis • Tasks for which it can be used: customer segmentation, forecasting sales, market basket analysis, identifying customer shopping behavior
Oracle Data Mining	<ul style="list-style-type: none"> • Embeds data mining techniques within the Oracle database • Provides means for building, testing, validating, managing and deploying data mining models inside the database environment • Supports the following data mining functions: classification, regression, attribute importance, anomaly detection, clustering, association models and feature extraction • Applications for which it can be used: customer segmentation, recommend next likely product, credit scoring, customer profitability or fraud detection
SAS Enterprise Miner	<ul style="list-style-type: none"> • Software application that provides data mining algorithms for creating predictive and descriptive models • Comprises an easy to use graphical user interface that helps with data preparation, summarization and exploration, as well as advanced predictive and descriptive modeling • Data mining techniques applied include: classification (decision trees, neural networks), clustering, regression, association rules • Tasks for which it can be applied: detect fraud, anticipate resource demands, increase acquisitions and curb customer attrition.
STATISTICA Data Miner	<ul style="list-style-type: none"> • Statistics and analytics software package that provides data analysis, data management, statistics, data mining and data

	<p>visualization functions</p> <ul style="list-style-type: none"> • Provides effective data pre-processing, cleaning, and filtering tools, along with tools for producing prediction models in various formats • Data mining methods available include: clustering, classification, regression, association and sequence analysis • Applications for which it can be used: customer segmentation, customer retention, credit scoring, market basket analysis or price optimization
--	---

Several researchers and organizations have conducted reviews and surveys of data mining tools. These examine and provide an overview of the behaviors, preferences and views of data mining, data science and analytic professionals.

One of these reports is Annual Rexer Analytics Data Miner Surveys, published by Rexer Analytics.

The selection of the primary commercial data mining tools in 2013, according to this report, is presented in Fig. 3 below:

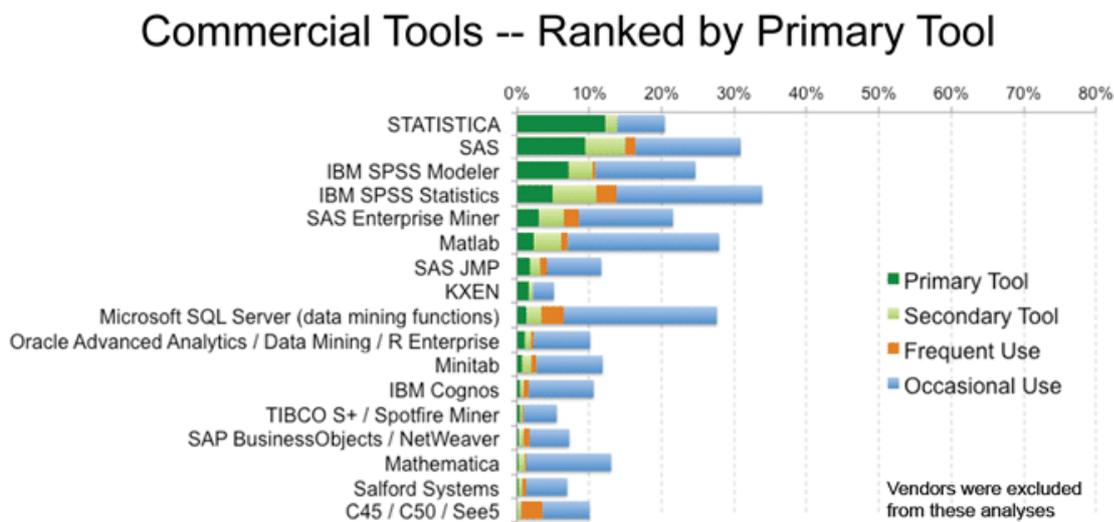


Fig. 3 Primary commercial data mining tools in 2013 [8]

The analysis of the above figure shows for the commercial data mining tools, which tools were considered by the respondents as primary tool, secondary tool, frequently or occasionally used tool. According to this survey, in 2013, the primary data mining tools used were STATISTICA Data Miner, IBM SPSS Modeler and SAS Enterprise Miner.

Each data mining tool analyzed has different features and can be used for various requirements.

5. Data mining solution for the business environment

Business is well-fit domain for applying data mining as it provides large volumes of data.

The main features of a data mining solution for the business environment, presented in Fig. 4 below, are:



Fig. 4 Main features of a data mining solution for business

- *Selecting the data*: identify the data sets used for a specific analysis and improve the initially selected data sets if required;
- *Preparing the data*: transform and clean the data so it is in the appropriate format for applying data mining techniques;
- *Choosing the data mining technique*: select the algorithm associated to a data mining technique that is suitable for the required analysis;
- *Configuring the settings for the data mining technique*: configure for the selected algorithm the necessary parameters;
- *Executing the data mining process*: execute the configured data mining process;
- *Viewing the results of the data mining process*: view the results generated by the execution of the data mining process.

The architecture of the proposed data mining solution, applied to the business environment, is illustrated in Fig. 5 below:

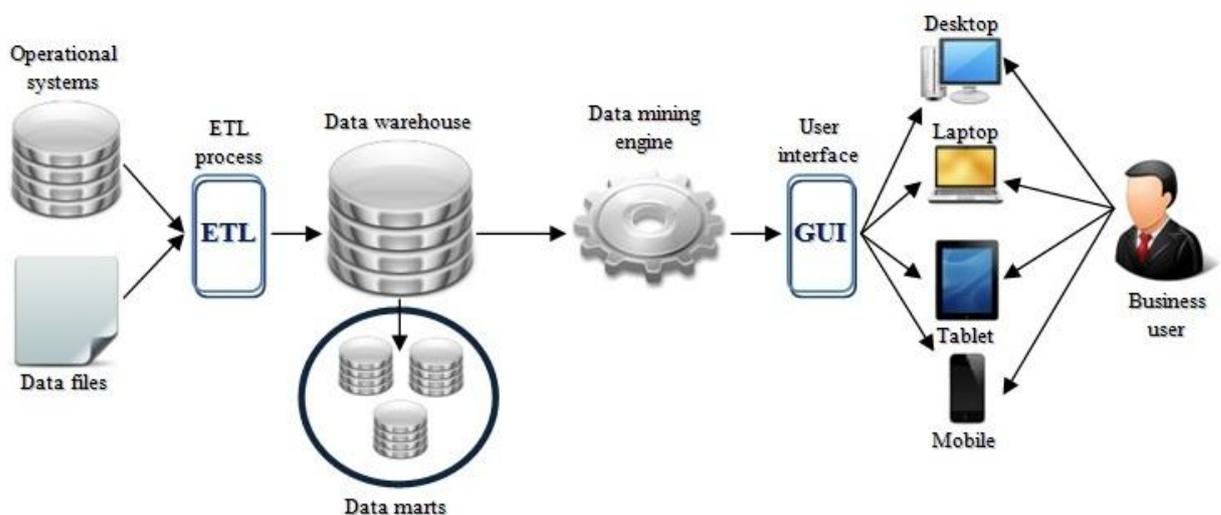


Fig. 5 Architecture of data mining solution for business

The data is extracted from the data sources, both operational systems and data files, and loaded through the ETL (Extract, Transform and Load) process to the data warehouse. The data warehouse can contain raw data – the data in a detailed format, as it has been extracted, summary data – data that has been aggregated and transformed – and metadata, data that provides information about the raw and summary data belonging to the data warehouse. The data belonging to the data warehouse can be organized in data marts.

The solution provides a data mining engine that may be used for obtaining advanced analysis. The solution has a graphical user interface that provides access to the main features of the solution, as presented in the article: select and prepare the data, choose data mining technique, configure the settings and execute the data mining process, view the results obtained.

The business user may access the GUI of the solution from various devices, like the desktop, laptop, tablet or mobile.

Using the functionalities described above the solution allows accessing and analyzing business related information in order to obtain valuable knowledge concerning the business.

Conclusions

Our current society needs data mining for improving many domains of human life. Business areas like retail, banking and insurance can use data mining methods to improve customer experiences, make optimal decisions, strengthen their market position and achieve competitive advantage.

There are various commercially available data mining tools to provide support for fulfilling these requirements.

The architecture proposed for the data mining solution for the business environment would improve the efficiency of a company, by providing

valuable decision-making knowledge to minimize operating costs and gain competitive advantage.

References

- [1] Usama Fayyad, Gregory Piatetsky-Shapiro and Padhraic Smyth, *From Data Mining to Knowledge Discovery in Databases*, AI Magazine, Vol. 17, Issue 3, 1996, ISSN 0738-4602, pp. 37-54.
- [2] Ion Lungu and Adela Bâra – “Improving Decision Support Systems with Data Mining Techniques”, “Advances in Data Mining Knowledge Discovery and Applications” – chapter 18, InTech Publisher, Croatia, 2012, ISBN 978-953-51-0748-4, pp. 397-418.
- [3] Oded Maimon and Lior Rokach, *Data Mining and Knowledge Discovery Handbook. Second Edition*, Springer Publishing Company, USA, 2010.
- [4] Gordon S. Linoff and Michael J. A. Berry, *Data Mining Techniques: for Marketing, Sales and Customer Relationship Management. Third Edition*, Wiley Publishing, USA, 2011.
- [5] Jiawei Han, Micheline Kamber and Jian Pei, *Data Mining: Concepts and Techniques. Third Edition*, Morgan Kaufmann Publishing, USA, 2011.
- [6] Vikas Jayasree and Rethnamoney Vijayalakshmi Siva Balan, *A Review on Data Mining in Banking Sector*, American Journal of Applied Sciences, Vol. 10, Issue 10, 2013, ISSN 1554-3641, pp. 1160-1165.
- [7] A. B. Devale and Dr. R. V. Kulkarni, *Applications of data mining techniques in life insurance*, International Journal of Data Mining & Knowledge Management Process, Vol.2, Issue 4, July 2012, ISSN 2230-9608, pp. 31-40.
- [8] Rexer Analytics, *Annual Rexer Analytics Data Miner Survey*, Commercial Tools – Ranked by Primary Tool, 2013, Available: <http://www.statsoft.com/Company/About-Us/Reviews/2013-Published-Reviews>.



Ruxandra PETRE graduated from the Faculty of Cybernetics, Statistics and Economic Informatics of the Bucharest University of Economic Studies in 2010. In 2012 she graduated the Business Support Databases Master program. Currently, she is a PhD candidate, coordinated by Professor Ion LUNGU in the field of Economic Informatics at the Bucharest University of Economic Studies. The title of her PhD thesis is “Information Solutions for Data Analysis”. Her scientific fields of interest include: Databases, Data Warehouses, Business Intelligence, Decision Support Systems and Data Mining.