5

Data Mining and Decision Support: An Integrative Approach

Rok Rupnik and Matjaž Kukar
University of Ljubljana, Faculty of Computer and Information Science
Slovenia

1. Introduction

Modern organizations use several types of application systems to facilitate knowledge discovery and decision support. Transactional application systems usually have sophisticated reports presenting data by using concepts like sorting, grouping and data aggregation. OLAP systems, also referred to as management information systems, use a data warehouse as a data source and represent a higher-level tool enabling decision support. In such a data warehouse, data are periodically extracted in an aggregated form from transactional information systems and other external sources by data warehouse tools. Both, transactional information systems and OLAP systems, are generally based on concepts of sorting, grouping and data aggregation, where with data aggregation one of the aggregating functions like sum, minimum, maximum, count and average are used. Both, transactional application systems reports and OLAP systems enable the presentation of different viewpoints on aggregated data in different dimensions, the latter however presenting more dimensions than the former (Bose & Sugumaran, 1999).

However, the advancement of strategies for information analyses, business prediction and knowledge extraction have lagged the corresponding developments in data storage and representation, especially in real world applications. Among the reasons for this, there are the inherent complexity of certain aspects of real world problems and the lack of data analysis expertise among business planners, which are compounded by confusing marketing literature produced by a few vendors about the capabilities of their analyses and prediction tools. These factors are combined to reduce the business planners' degree of belief in many of these tools, which in turn leads to these tools being given less importance in key decisions, hence causing the IT vendors to focus less on the development and implementation of the cutting edge techniques.

Traditionally, statistical and OLAP tools have been used for an advanced data analysis. It is often assumed that the business planners would know the specific question to ask, or the exact definition of the problem that they want to solve. Both methods follow what is in essence a deductive approach (Jsr73, 2004), which has several drawbacks. They put a significant strain on the data analyst, who must take care of inventing a hypothesis and storing it in an appropriate way (Hirji, 2001). They lack algorithmic approach and depend on the analysts' insight, coincidence or even luck for acquiring the most valuable information, trends and patterns from data. And finally, even for the best analyst there is a limitation to a number of attributes he can simultaneously consider in order to acquire

Source: Decision Support Systems, Book edited by: Chiang S. Jao,
ISBN 978-953-7619-64-0, pp. 406, January 2010, INTECH, Croatia, downloaded from SCIYO.COM

accurate and valuable information, trends and patterns (Goebel & Gruenwald, 1999). It seems that with the increase in data volume, traditional manual data analysis has become insufficient, and methods for efficient computer-based analyses indispensable. From this need, a new interdisciplinary field of data mining arose. Data mining encompasses statistical, pattern recognition, and machine learning tools to support the analysis of data and discovery of principles that lie within data given.

In the past decades several data mining methods have emerged, showing high potentials for knowledge discovery from data and decision support. Performing analysis through data mining follows an inductive approach of analyzing data where machine learning algorithms are applied to extract non-obvious knowledge from data (Jsr73, 2004). Data mining reduces or even eliminates the above mentioned disadvantages. As opposed to classical data analysis techniques, data mining strategies often take a slightly different view, i.e. the nature of the data itself could dictate the problem definition and lead to discovery of previously unknown but interesting patterns in diverse business applications, for example sales forecasting during promotion events, inventory optimization, and customer profiling and segmentation.

Data mining methods also extend the possibilities of discovering information, trends and patterns by using richer model representations (e.g. decision rules, trees, tables , ...) than the usual statistical methods, and are therefore well-suited for making the results more comprehensible to the non-technically oriented business users. It may well be that by the introduction of data mining to information systems the knowledge discovery process and decision process will move to a higher quality level.

The mission of information systems is, among other things, to facilitate decision support and knowledge discovery. Both, the decision process and the knowledge discovery process are dependent on each other. Knowledge discovery, on one hand, enables accumulation of knowledge and as a result facilitates better decision process. On the other hand, decisions set rules and directions which influence objectives for knowledge discovery. The use of data mining within information systems consequently means the semantic integration of data mining and decision support.

1.1 Motivation

The motivation for our pioneering work in integration of decision support system with data mining methods originates from a real-world problem. Recently we have performed an extensive CRM survey for a leading local GSM operator.

One of the aims of the survey was to explore and demonstrate various approaches and methods for the area of analytical CRM. The survey clearly revealed the benefits of the use of data mining for analytical CRM. The survey was performed between 2002 and 2003 by the authors of the chapter, who had primary roles in the survey project.

As the survey progressed, it turned out that immense quantities of raw data had been collected and needed to be assessed. Immediately after the survey had been conducted, the development project for data mining application system was initiated and managed by the group executing survey. The application system is called DMDSS (Data Mining Decision Support System). DMDSS application system will be introduced later on in the chapter.

1.2 Structure of the chapter

The chapter is organized as follows. In Section 2 we are introducing the basic concepts, underlying ideas of data mining in general and in particular. We are focusing on data

mining subfields and particular methods that were used in our study. We are going to outline crucial similarities and differences between data mining on one side and on-line analytical processing and statistical approaches on the other side.
In Section 3 we are introducing motivation for the use of data mining within information systems, and highlight some historical examples of case studies. We are presenting historical generations of data mining, data mining standards, data mining process model (CRISP-DM) and provide motivation for its use in practice.
In Section 4 we are going to proceed to our case study of using data mining in a decision support system developed for a leading local GSM operator. We are describing the development process, review some practical considerations, introduce functionalities of DMDSS and present end-users' experiences after one year of practical use. At the end we are introducing the semantic contribution of the use of DMDSS in the information system.
In Section 5 we are going to summarize our work and provide some conclusions as well as directions for future work.

2. An introduction to data mining

"*Now that we have gathered so much data, what do we do with it?*"

This is the famous opening statement of the editorial by Usama Fayyad and Ramasamy Uthurusamy in the Communications of the ACM, Special issue on Data Mining (Fayyad & Uthurusamy, 1996). Recently, many statements of this kind have appeared in journals, conference proceedings, and other materials that deal with data analysis, knowledge discovery, and machine learning. They all express concern about how to "make sense" from the large volumes of data being generated and stored in almost all fields of human activity.
Especially in the last few years, the digital revolution has provided relatively inexpensive and available means to collect and store the data. The increase in data volume causes greater difficulties in extracting useful information for decision support. The traditional manual data analysis has become insufficient, and methods for efficient computer-based analysis indispensable. From this need, a new interdisciplinary field of data mining was born. Data mining encompasses statistical, pattern recognition, and machine learning tools to support the analysis of data and discovery of principles that lie within the data.
Results of computer-based analysis have to be communicated to humans in an understandable way. In this respect, the analysis tools have to deliver transparent results and most often facilitate human intervention in the analysis process. A good example of such methods are symbolic machine learning algorithms which, as a result of data analysis, aim to derive a symbolic model (e.g., a decision tree or a set of rules) of preferably low complexity but high transparency and accuracy. Being in the core of data mining, the interest and research efforts in machine learning have been largely increased.
Data mining is about automated extraction of hidden information from large databases. They consist of digitised information which is easy to capture and fairly inexpensive to store. But why do people store so much data? Besides the fact that it is easy and convenient to do so, people store data because they think some valuable assets are implicitly coded within it. In scientific endeavours, data represents observations carefully collected about some phenomenon under study. In business, data captures information about critical markets, competitors, and customers. In manufacturing, data captures performance and optimisation opportunities, as well as the keys to improving processes and troubleshooting problems.

Raw data is rarely of direct benefit (Witten & Frank, 2000). Its true value is predicated on the ability to extract information useful for decision support or exploration and understanding the phenomena governing the data source. Traditionally, the analysis was strictly a manual process. One or more analysts would become familiar with the data and—with the help of statistical techniques—provide summaries and generate reports. In fact, the analysts acted as sophisticated query processors. However, such an approach is rapidly breaking down as the quantity of data grows and the number of dimensions increases. Who could be expected to "understand" millions of cases, each having hundreds of fields (attributes)? Further complicating this situation, the amount of data is growing so fast that manual analysis (even if possible) simply cannot keep pace.

The allure of data mining is that it promises to improve the communication between users and their large volumes of data and allows them to ask of the data complex questions such as: "What has been going on?" or "What are the characteristics of our best customers?" The answer to the first question can be provided by the data warehouse and multidimensional database technology (OLAP) that allow the user to easily navigate and visualize the data. The answer to the second question can be provided by data mining tools built on variety of machine learning algorithms: decision trees and rules, neural networks, nearest neighbour, support vector machines, and many others (Chen et al., 1996).

Data mining takes advantage of advances in the fields of artificial intelligence (AI) and statistics. The challenges in data mining and learning from data have led to a revolution in the statistical sciences (Hasti & Tibisharani, 2002). Both disciplines have been working on problems of pattern recognition and classification. Both communities have made great contributions to the understanding and application of different paradigms, such as neural nets and decision trees.

Since statistical discovery is essentially a hypothesis-driven deductive process (Hirji, 2001), an analyst generates a series of hypotheses for patterns and relationships, and uses statistical tests against the data to verify or disprove them. But what happens when the number of variables being analyzed is in the dozens or even hundreds? It becomes much more difficult and time-consuming to find a good hypothesis (let alone be confident that there is not a better explanation than the one found), and analyze the database with statistical tools to verify or disprove it.

Data mining is different from statistical discovery because rather than to verify hypothetical patterns, it uses the data itself to uncover such patterns. It is essentially a discovery-driven inductive process, where a data mining tool is used to formulate (induce) hypotheses from data completely by itself or with moderate guidance from the analyst (Glymour et al., 1996).

Data mining does not replace traditional statistical techniques. On the contrary; it is an extension of statistical methods that is in part the result of a major change in the statistics community. The development of most statistical techniques was, until recently, based on elegant theory and analytical methods that worked quite well on the modest amounts of data being analyzed. This is especially the case in sensitive application areas such as finance or medicine (Kukar et al., 1999; Kononenko, 2001) where useful results are required regardless of the amount of data.

The increased power of computers and their lower cost, coupled with the need to analyze enormous data sets, have allowed the development of new techniques based on a brute-force exploration of possible solutions.

New techniques include relatively recent algorithms like neural nets, decision rules and trees, and new approaches to older algorithms such as discriminant analysis. By virtue of

bringing to bear the increased computer power on the huge volumes of available data, these techniques can approximate almost any functional form or interaction on their own, as well as provide additional information on the functional solution (Kukar, 2003). Traditional statistical techniques rely on the modeler to specify the functional form and interactions.
The key point is that data mining is the application of these and other AI and statistical techniques to common business problems in a fashion that makes these techniques available to the skilled knowledge worker as well as the trained statistics professional. Data mining is a tool for increasing the productivity of people trying to build predictive models.

3. Introduction of data mining to information systems

Modern organizations use several types of application systems to facilitate knowledge discovery and decision support. Transactional application systems have sophisticated reports presenting data using concepts like sorting, grouping and data aggregation. Data is aggregated on the group level using one of the following aggregating functions: sum, minimum, maximum, count, average, standard deviation and variance. For example: a report can be grouped by customers by regions, showing average profits won in descending order. Such reports represent the basic level of facilitating decision support on tactical level.
OLAP systems, also referred to as management information systems, use data warehouse as data source and represent higher level tool enabling decision support on tactical level. In a data warehouse data are periodically extracted in aggregated form from transactional information systems and other external sources by data warehouse tools. A data warehouse is typically a multidimensional structure where dimensions represent attributes. OLAP systems enable drill-down concept, i.e. digging through a data warehouse from several viewpoints to acquire the information the decision maker is interested in (Bose & Sugumaran, 1999).
Both, transactional information systems reports and OLAP systems, in general base on concepts of sorting, grouping and data aggregation, where with data aggregation one of the aforementioned aggregating functions are used. They both enable the presenting of different viewpoints on aggregated data by different dimensions. Through the use of sorting they also enable the observing of better/worse relations among elements within individual dimension, i.e. attribute. For example: the form in an OLAP system contains a three-dimensional view showing the profit won by products by regions by departments. If the profit is sorted in descending order, the analyst can observe better/worse relations within the profit won by products within region and department.
Transactional information systems reports and OLAP systems both support the knowledge discovery and analysis process, where the analysts are supposed to look for information, trends and patterns. They do it by running numerous reports with several parameter sets and by viewing OLAP forms swapping dimensions and drilling-down through them (Goebel & Gruenwald, 1999). Performing analysis through OLAP and reports running follows a deductive approach of analyzing data (Jsr73, 2004). Such an approach has the following disadvantages:

- This way analysts can discover information, trends and patterns, but they must take care of storing it themselves to a shared source in a form and structure that will enable the knowledge assimilation and the exploiting of knowledge (Hirji, 2001),

- It lacks algorithm-based approach and depends on coincidence or even luck of choosing the right parameter set or the right dimensions to acquire the most valuable information, trends and patterns,
- There is a limitation to a number of attributes a human being can simultaneously consider in order to acquire accurate and valuable information, trends and patterns (Goebel & Gruenwald, 1999).

The introduction of data mining in previous section indicates high potentials of the use of data mining to facilitate knowledge discovery and decision support. Performing analysis through data mining follows an inductive approach of analyzing data where machine learning algorithms are applied to extract non-obvious knowledge from data (Jsr73, 2004). Data mining, on one hand, reduces or even eliminates before mentioned disadvantages, because:

- It enables the creation of models explaining trends and patterns, which can be stored in a standardized form to a shared source,
- It is an algorithm-based approach, because the models are acquired by data mining methods and algorithms,
- The majority of data mining algorithms are capable of creating models based on a large number of attributes.

3.1 Some examples of the use of data mining in information systems

There are a number of examples of successful application of data mining in information systems. In this section we are introducing some of them. Web usage mining is the application of data mining methods to discover patterns from Web data. Web Data can be classified into the following types: content, structure, usage and user profile. The purpose of Web usage mining is to understand better how users use e-commerce applications in order to improve them. The main application areas of Web mining are personalization, system improvement, site modification, business intelligence and usage characterization (Srivastava et al., 2000). Organizations conducting electronic commerce can, without any doubt, benefit from the use of data mining (Kohavi et al., 2002).

Lee and Stolfo (1998) have developed general and systematic methods for network intrusion detection. They use data mining techniques to discover consistent and useful patterns of system features that describe program and user behaviour, and use the set of relevant system features to compute inductively learned classifiers that can recognize anomalies and known intrusions. The discovered patterns help guide the audit data gathering process and facilitate feature selection (Lee & Stolfo, 1998).

A customer retention analysis represents an important type of analysis in the area of analytical CRM (Customer Relationship Management). It represents the area where data mining can be effectively used. A customer retention analysis is extremely important for sales and service related businesses, because it costs significantly more money to acquire a new customer than to retain existing customers. The key application in the area of customer retention is the detection of potential defectors. The possibility to predict defectors is often the key to the eventual retention of the customer (Ng & Liu, 2000; Han et al., 2002).

The education domain offers a fertile ground for many interesting data mining applications. Selecting the right students for a particular course is one of the application areas within the education domain. The aim of the application mentioned is to help students select courses in which they have high prospects to perform well (Ma et al., 2000). There are plenty of other

application areas within education domain appropriate for analysis purposes. For example: predictive model for excellent, good and bad students using classification method.
Several authors have indicated insurance as one of the areas with the highest potentials in the use of data mining (Grossman et al., 2002). Fraud detection is the key application area for data mining in insurance companies. A U.S. health insurance company, for example, used data mining methods to detect submitting of false bills. The analysis identified several geographical areas where claims exceeded the norm and the investigation confirmed and detected physicians who were submitting false bills (Furlan & Bajec, 2008).
We believe that in mobile telecommunication industry there is also a high potential in the use of data mining methods. Besides customer retention there are a number of interesting data mining application areas within analytical CRM:

- **Customer segmentation**: the idea of customer segmentation is to acquire typical customer groups in order to perform group targeted marketing.
- **Subsidized mobile phone value analysis**: subsidized mobile phones represent a substantial investment by mobile operator. The idea of the analysis is to acquire a classification model for the customers spending significantly more money for GSM services after purchasing subsidized mobile phone.
- **Various CDR (call data record) analyses**: in a CDR database there are billions of records that represent a good foundation for various areas of analysis, which among others represent a basis for tariff plans adjustments and wireless network planning.
- **Various association analyses**: there are several options for a valuable analysis of relations. For example: the observation of consequences of the level of the use of services after switching a tariff plan.

There are many other areas where data mining was successfully used (Fayyad et al., 1996; Han et al., 2002; Kohavi et al., 2002). Generally we can say that data mining can be effectively used when sufficient amount of data described with sufficient amount of high quality attributes is available.

3.2 Generations of data mining

Examples introduced in previous sections show that companies can use data mining within their information systems in various areas with different objectives. The use of data mining is increasing, but has still not reached the level appropriate to the potential benefits of its use (Kohavi & Sahami, 2000). One of the reasons for that is with no doubt the lack of awareness and understanding of potentials of data mining. We are presenting other reasons in the course of introducing two generations of data mining, which we introduce as two different approaches.

3.2.1 First generation: data mining software tool approach

Data mining today and in the past has been typically used through ad hoc data mining projects (Goebel & Gruenwald, 1999; Kohavi & Sahami, 2000; Holsheimer, 1999). Ad hoc data mining projects are initiated by a particular objective on a chosen area which means defining of the domain. They are performed using data mining software tools. It is the first generation of data mining (Holsheimer, 1999).
Data mining software tools require a significant expertise in data mining methods, modelling methods, databases and/or statistics (Kohavi & Sahami, 2000). They usually operate separately from the data source, requiring a significant amount of additional time

spent with data export from various sources, data import, pre-processing (extraction, filtering, manipulation), post-processing (validation, reporting) and data transformation (Holsheimer, 1999; Goebel & Gruenwald, 1999). The result of a project is usually a report explaining the models acquired during the project using various data mining methods.
Data mining software tool approach has a disadvantage in a number of various experts needed to collaborate in a project and in transferability of results and models (Srivastava et al., 2000; Hirji, 2001). The latter indicates that results and models acquired by the project can be used for reporting, but cannot be directly utilized in other application systems.
The disadvantages mentioned can be explored by the disadvantages of data mining software tools, which are implicitly also disadvantages of data mining software tool approach. Data mining software tools are very complex, they often offer a variety of methods that a user must understand in order to use them effectively. Some of the tools do not enable on-line access to a database or they enable on-line access only to a few database systems. On the other hand, some tools do not allow an access to any database and in this case data must first be extracted from a database to a file before used by a tool. Only a few data mining tools support pre-processing activities (Goebel & Gruenwald, 1999; Kohavi & Sahami, 2000).
In our opinion some of the disadvantages mentioned are in a way also advantages. Complexity of tools and a number of various experts needed are in most cases viewed as disadvantages due to the fact that there are not many experts providing the cutting edge of data mining methods and software tools. The involvement of various experts and a data mining expert providing expertise in data mining methods and in complex tool can only positively contribute to the results of a data mining project.

3.2.2 Second generation: a data mining application system approach

The data mining software tool approach has revealed some disadvantages, which point to the following demands for a different approach:

- The end users of the models and results acquired by data mining projects are business users. For that reason we need applications which will enable them to use data mining models effectively (Kohavi & Sahami, 2000; Goebel & Gruenwald, 1999).
- Business users are not interested in using advanced powerful software tools, but only in getting clear and fast answers by using simple-to-use applications (Goebel & Gruenwald, 1999; Holsheimer, 1999; Bose & Sugumaran, 1999; Fayyad &Uthurusamy, 2002).
- In order to achieve the highest level of the use of data mining models and results it must be possible to deploy them to other business applications in order to use them there (Holsheimer, 1999; Geist, 2002).
- The models and results acquired are dependent on data which is not stationary, but is constantly changing and evolving. For that reason ad hoc projects and a data mining software tool approach need to be enhanced by repeating the same model creation (analysis) process in periodic time intervals or at particular milestones (Goebel & Gruenwald, 1999).

The list of demands indicates the characteristics of application system approach for the use of data mining. It is an approach which focuses on business users, enabling them to view data mining models and results in their business domains. Models and results are presented in a user-understandable manner by means of a user friendly and intuitive GUI using standard and graphical presentation techniques (Aggarwal, 2002). Business users can focus

on specific business problems covered by domains with the possibility of repeated analysis in periodic time intervals or at particular milestones. Through the use of data mining application systems approach, data mining becomes better integrated in business environments (Goebel & Gruenwald, 1999; Holsheimer, 1999; Kohavi & Sahami, 2000; Bayardo & Gehrke, 2001; Fayyad & Uthurusamy, 2002).
The data mining application system approach is enabled by application systems which use data mining methods. Those application systems can be divided into two categories. The first category are application systems which support the whole knowledge discovery process, where one set of functionalities is used by data mining experts and the rest of the functionalities by business users as described before. We call them data mining application systems. The second category is other business application systems which can utilize data mining models for various purposes. They can either directly access data mining models, or data mining models can be deployed to them. An example of data mining application will be introduced later on in the chapter. The discussion will reflect advantages and disadvantages of data mining software tool approach.

3.3 Data mining standards

Data mining standards undoubtedly represent an important issue for data mining application systems approach and data mining application systems (Holsheimer, 1999). Employing common standards simplifies the development of data mining application systems and business application systems utilizing data mining models (Grossman et al., 2002; Grossman, 2003). With the maturity of data mining standards, a variety of standards-based data mining applications and data mining platforms can be much easily developed (Grossman, 2003). Other fields such as data grids, web services and the semantic web have also developed standards relevant to knowledge discovery (2003; Chu, 2003; Kumar & Kantardzic, 2003). These new standards have the potential for further changing the way the data mining is used (Grossman, 2003).
A considerable effort in the area of data mining standards has already been done within the data mining community. Established and emerging data mining standards address several aspects of data mining (Grossman et al., 2002):

- Models: for representing data mining models,
- Attributes: for representing the cleaning, transforming and aggregating of the attributes used as input for model creation,
- Settings: for representing the algorithm parameters which affect the model creation,
- Process: for creating, deploying and utilizing the models,
- APIs: for unified access to all methods enabling models creating, deploying and utilizing,
- Remote and distributed data: For analyzing and mining remote and distributed data.

In the following subsections we introduce the most important data mining standards.

3.3.1 PMML

The Predictive Model Markup Language (PMML) is an XML standard being developed by the Data Mining Group, a vendor led consortium established to develop data mining standards (Grossman et al., 2002; Grossman, 2003; Clifton & Thuraisingham, 2001; Dmg). It consists of the following components: data dictionary, mining schema, transformation dictionary, model statistics and models. PMML describes data mining and statistical models

in addition to some of the operations required for data cleaning and transforming prior to modelling. The aim of PMML is to provide infrastructure for an application to create a model and another application to utilize it (Grossman et al., 2002). PMML has been evolving from version 1.0 to version 2.1., changes and improvements for version 3.0 are being considered as well (Meyer, 2003). PMML directly covers the aspects of models, attributes and settings. Implicitly it also covers the aspect of API, because it provides the standard for infrastructure for manipulating with models: creating, deploying and utilizing.

3.3.2 JDM: the standardised data mining API

The standardized data mining API represents the most important issue for data mining application systems approach having the following advantages:

- Data mining algorithms are not coded by each team of application developers, but by teams which are specialized on data mining algorithms. Consequently, more reliable and efficient algorithms are developed,
- The possibility to leverage data mining functionality using standard API shared by all application systems within information systems reduces risk and cost,
- Standardized API facilitates API of one vendor to be replaced with API of another vendor.

Java technology and scalable J2EE architecture facilitate integration of various application systems within information system. For that reason many business application systems have been developed on J2EE platform in recent time. Consequently Java is probably the best option for standard data mining API, because it enables the integration of data mining application systems with other business applications within information systems (Hornick, 2003). Java based data mining API in a very effective way enables the implementation of data mining application systems approach and the development of data mining application systems and the integration of data mining in other business application systems.

JDM (Java Data Mining) specification has reached final release status in 2004 (JsrHp73). JDM specifies a pure Java API to facilitate development of Java-based data mining application systems and business application systems utilizing data mining models. As existing data mining APIs are vendor-proprietary, this is the first standardised API for data mining. The JDM expert group consists of representatives of several key software companies (including data mining pioneers IBM, SPSS and Oracle), what gives a certain guarantee for an exploitable standard. Detailed introduction of JDM can be found in (Jsr73, 2004).

3.3.3 CRISP-DM: data mining process model

A data mining process model defines the approach for the use of data mining, i.e. phases, activities and tasks that have to be performed. Data mining represents a rather complex and specialised field. A generic and standardized approach is needed for the use of data mining in order to help organizations use the data mining.

CRISP-DM (CRoss-Industry Standard Process for Data Mining) is a non-proprietary, documented and freely available data mining process model. It was developed by the industry leaders and the collaboration of experienced data mining users, data mining software tool providers and data mining service providers. CRISP-DM is an industry-, tool-, and application-neutral model created in 1996 (Shearer, 2000; Clifton & Thuraisingham, 2001; Grossman et al., 2002). Special Interest Group (CRISP-DM SIG) was formed in order to further develop and refine CRISP-DM process model to service the data mining community

well. CRISP-DM version 1.0 was presented in 2000 and it is being accepted by business users (Shearer, 2000).
CRISP-DM process model breaks down the life cycle of data mining project into the following six phases which all include a variety of tasks (Shearer, 2000; Clifton & Thuraisingham, 2001):

- **Business understanding**: focuses on understanding the project objectives form business perspective and transforming it into a data mining problem (domain) definition. At the end of the phase the project plan is produced.
- **Data understanding**: starts with an initial data collection and proceeds with activities in order to get familiar with data, to discover first insights into the data and to identify data quality problems.
- **Data preparation**: covers all activities to construct the final data set from the initial raw data including selection of data, cleaning of data, the construction of data, the integration of data and the formatting of data.
- **Modelling**: covers the creation of various data mining models. The phase starts with the selection of data mining methods, proceeds with the creation of data mining models and finishes with the assessment of models. Some data mining methods have specific requirements on the form of data and to step back to data preparation phase is often necessary.
- **Evaluation**: evaluates the data mining models created in the modelling phase. The aim of model evaluation is to confirm that the models are of high quality to achieve the business objectives.
- **Deployment**: covers the activities to organize knowledge gained through data mining models and present it in a way users can use it within decision making.

There are some other data mining process models found in the literature. They use slightly different terminology, but they are semantically equivalent to CRISP-DM (Goebel & Gruenwald, 1999; Li et al., 2002).

4. DMDSS – Data Mining Decision Support System for GSM operator

We have developed a data mining application system for a GSM operator who was the first on the market. In the following part of the chapter it will be called simply a GSM operator. One of their measures to keep prevailing position on the market was the initiation of the survey of CRM implementation.
One of the aims of the survey was to explore and demonstrate various approaches and methods for the area of analytical CRM. The survey clearly revealed the benefits of the use of data mining for analytical CRM. The important statement of the survey was that our GSM operator intolerably needs data mining for performing analysis for CRM purposes. It was stated that the application system approach is more suitable for the introduction of data mining. That statement was outlined after the research introduced in the previous section was conducted. The main reason for choosing the application system approach was the fact that CRM in our GSM operator represent a rather dynamic environment with continual need for repeated analyses based on data mining methods. The results of the survey were several reports and some prototypes. The survey was performed between 2002 and 2003 by the authors of the chapter, who had primary roles at the survey project.
Right after the survey, the development project for data mining application system was initiated and managed by the group executing the survey. The application system is called

DMDSS (Data Mining Decision Support System). DMDSS application system will be introduced in the following part of the chapter.

4.1 Related work

Several data mining application systems have been developed in recent years and introduced in scientific literature. In this section we introduce some of them.

Geist (2002) introduced a framework for data mining application system. He proposes a three view architecture consisting of process view, model view and data view. The process view covers the user interaction with the application system supporting management and controlling of data mining process. It supports the analysis process by supporting different presentation modes. The process view uses the methods offered by the data view and the model view. The data view covers raw data sets providing methods for manipulating the data objects and describing structure of the data. A relational data model is proposed as architecture for a data view. The model view consists of a set of data mining models and methods manipulating models. It supports model creation and the access to all information about the data mining model.

Bose and Sugumaran (1999) introduced Intelligent Data Miner (IDM) application system in his paper. IDM is a Web-based application system intended to provide organization-wide decision support capability for business users. Besides data mining it also supports some other function categories to enable decision support: data inquiry, data interpretation and multidimensional analysis. In the data mining part it supports the creation of models of the following data mining methods: association rules, clustering and classification. Through the use of various visualization methods it supports the presentation of data mining models. On the top-level it consists of the following five agents: user interface agent, IDM coordinator agent, data mining agent, data-set agent and report/visualization agent. The user interface agent provides interface for the user to interact with IDM to perform analysis. It is responsible for receiving user specifications, inputs, commands and delivering results. The IDM coordinator agent is responsible for coordinating tasks between the user interface agent and other three of before mentioned agents. Based on the user specifications, input and commands, it identifies tasks that need to be done, define the task sequence and delegates them to corresponding agents. It also synthesises and generates the final result. The data-set agent is responsible for communication with data sources. It provides interface to data warehouses, data marts and databases. The data mining agent is responsible for creating and manipulating of data mining models. It performs data cleansing and data preparation, provides necessary parameters for data mining algorithms and creates data mining models through executing data mining algorithms. The report/visualization agent is responsible for generation of the final report to the user. It assimilates the results from data mining agent, generates a report based on predefined templates and performs output customization.

Holsheimer introduced the Data Surveyor application system in his papers (Holsheimer, 1999; Holsheimer et al., 1995). He did not emphasize the functionalities it offers to the user. Instead, he put emphasis on the implementation of data mining methods and the interaction between Data Surveyor and database systems. Author describes Data Surveyor as a system designed for the discovery of rules. It is a 3-tier application system providing customized GUI for two organization roles: the data analyst and the end user. It enables the use of data in several RDBMS systems. Important characteristic of Data Surveyor is the ability to store data mining models in RDBMS system.

Heindrichs and Lim (2003) have done research on the impact of the use of web-based data mining tools and business models on strategic performance capabilities. His paper reveals web-based data mining tools to be a synonym for data mining application system. The author states that the main disadvantage of data mining software tool approach is the fact that it provides results on a request basis on static and potentially outdated data. He emphasizes the importance of the data mining application system approach, because it provides ease-of-use and results on real-time data. The author also discusses the importance of data mining application systems through arguing that sustaining a competitive advantage in the companies demands a combination of the following three prerequisites: skilled and capable people, organizational culture focused on learning, and the use of leading-edge information technology tools for effective knowledge management. Data mining application systems with no doubt contribute to the latter. In the paper the author also introduces the empirical test which proves positive effect on dependent variable "Strategic performance capabilities" by independent variables "Web-based data mining tools" and "Business models".

4.2 Pre-development activities

The development of DMDSS started with several pre-development activities. We are going to introduce them in the following sections.

4.2.1 Platform selection

The first step within the pre-development activities was the platform selection. Platform selection was highly influenced by two important factors. The first factor was the dominant presence of Oracle platform in our GSM operator. More than 80% of data needed for data mining was available in Oracle databases. The second one was the fact that Oracle RDBMS 9i introduced ODM (Oracle Data Mining) option. ODM has two important components. The first component is a data mining engine (DME), which provides the infrastructure that offers a set of data mining services to data mining API (JSR73, 2004). The second component is Java-based data mining API (ODM API), which enables access to services provided by DME.

Before finally accepting Oracle 9i and ODM, an evaluation sub-project was initiated. The aim of the project was to evaluate ODM, i.e. to verify the quality of its algorithms and results. It was the first version of ODM and evaluation was simply necessary to reduce the risk of using an immature product. The evaluation was performed through recreating data mining models on domains of some past projects which were performed by our research group using data mining software tools (Kukar, 2003; Kukar et al., 1999; Kononenko, 2001). The models acquired by past projects and the results acquired by ODM were compared and evaluation gave positive results for the verification of ODM. Another advantage of Oracle 9i is the security issue. As opposed to many other data mining platforms, in case of Oracle 9i data mining, data does not leave the database. Data mining models and their rules are stored in the database, which means that database security provides the control over access to data mining data, i.e. models and rules.

The introduced factors and the result of the sub-project implied the selection of Oracle 9i RDBMS. In order to develop DMDSS in J2EE architecture the JDeveloper development platform and Oracle OC4J were chosen.

4.2.2 Functional and other demands for DMDSS

The analysis of functional and other demands for DMDSS was done simultaneously with the design of data mining process model for DMDSS. Both activities are extremely interrelated, because the process model implies the functionality of an application to a great extent. The design of data mining process model for DMDSS is introduced in the next section.

It turned out that DMDSS directly or indirectly needs three roles: a data administrator, a data mining administrator and a business user. A data mining administrator role should be granted only to users with advanced or at least above-average knowledge of data mining methods and concepts. Business users are business analysts responsible for performing analysis in various business areas.

The analysis of functional and other demands led to the following important conclusions:

- The roles of the data mining administrator and the business user must be supported. Data administrator role and data preparation phase will not be supported by DMDSS, they should be supported by other tools.
- The access to the modules and functionalities should be dependent on user's role. This would prevent business users from using functionalities which demand advanced knowledge of data mining.
- The data mining administrator should have the possibility to create, evaluate and delete models. A set of model statuses should be defined in order to enable the administrator to make only good and useful models available for the business users.
- The data mining administrator should have the possibility to comment on the models and insert them in the database. Business users should have the possibility to see them, which would help them understand and interpret the models better.
- There should be various visualization and representation techniques available in order to enable various methods for model presentation for the business users.
- Before the training of business users there should be a data mining tutorial organized, where they could learn the concepts of data mining, which would enable them to use and truly exploit DMDSS.
- The key issue for the success of DMDSS is to define its functionalities in the way that will enable data mining administrator create and evaluate models. On the other hand, business users should be able to use it effectively with as little data mining knowledge as possible.

4.2.3 Data mining process model for DMDSS

The key pre-development activity was to determine the data mining process model for DMDSS, which would be appropriate for analysts in marketing department of our GSM operator. According to the level of their knowledge of data mining concepts it was obvious that DMDSS process model should enable analysts incorporate it in their decision process.

The analysis of CRISP-DM and other previously introduced data mining process models revealed that they are more appropriate for ad-hoc projects and a data mining software tool approach than for a data mining application system approach. The consequence was that none of them could be directly used for DMDSS and data mining application system approach. The analysis of data mining process models confirmed CRISP-DM as the most appropriate process model.

CRISP-DM was adapted to the needs of DMDSS as a three stage model where the last stage represents the final process model gained through first two stages. The first stage was the execution of business understanding phase, where the aim was to discover the domains with continual need for repeated analysis based on data mining methods. They are referred to as the areas of analysis.
The second stage was the execution of a data mining project for each area of analysis using a data mining software tool approach. The second stage was actually performed through development process of DMDSS, where multiple iterations of all CRISP-DM project phases were supported by iterations of development process and increments of DMDSS. The development process is introduced later on in the chapter.
The aim of executing multiple iterations of all CRISP-DM phases for every project was to achieve improvements in the areas of data preparation and to do the fine-tuning of data mining algorithms used in ODM API through finding proper parameter values for algorithms. Data sets were re-created automatically every night, based on the current state of the data warehouse and transactional databases. After the re-creation of data sets, data mining models were created and evaluated. It was essential to do iterations over longer period of time in order to implement automated procedures for data preparation and monitor the level of changes in data sets and data mining models acquired. One of the demands for DMDSS was the ability for daily creation of models for every area of analysis and for that reason the degree of changes in data sets and data mining models acquired were monitored.
The third stage represents the production phase of DMDSS and the final process model. Multiple iterations performed in the second stage assure the stability of data preparation phase and proper parameter value sets for data mining algorithms for modelling phase. Modelling and evaluation are performed by data mining administrator and deployment by business users. Some other details of DMDSS regarding process model will be introduced later on in the chapter.

4.3 Development of DMDSS

DMDSS was developed by using several diagramming techniques. UML use case diagrams and class diagrams were used for the process modelling of DMDSS. Entity relationship diagrams were used for data modelling. For several reasons we decided to use iterative incremental process model. As already mentioned, one of the reasons for multiple iterations was to achieve improvements and stability in the areas of data preparation and modelling. Iterations were also needed for incremental changes and improvements of functionalities of DMDSS. After the iteration had been finished, the functional testing was performed. Only then the analysis of functionalities were conducted done and based on that, the list of changes and improvements. The list of changes and improvements was used as the list of demands for the next iteration of development.
During development process we developed our data mining API (DMDSS API) based on ODM API. We did it because the interface of ODM API was badly documented and rather inconsistent, especially method naming. A considerable level of knowledge of Java and data mining algorithms was needed to understand ODM API interface and fully exploit it. For that reason we decided to develop DMDSS API, which would have the intuitive interface and method naming. The structure of DMDSS API interface was constructed in order to obscure the frequently changing details of the ODM API implementation, as well as to

provide a consistent platform for both supervised and unsupervised data mining. DMDSS API implied the division of development team into two groups: a team developing DMDSS GUI and a team developing DMDSS API. Such a division of the development team was efficient, because the development process could be carried out consequently to a certain extent and developers could be grouped according to their areas of specialization and skills.

4.4 The introduction of DMDSS

In this section we are going to introduce DMDSS. First we are going to introduce its role-aware architecture and roles using DMDSS. Then we are going to introduce concepts of the use and functionalities of DMDSS through some example forms for data mining administrator and business user. The introduction of concepts of use and functionalities is done for classification data mining method supported by DMDSS. We are going to finish the introduction of DMDSS by presenting the experience of the use of DMDSS in our GSM operator.

4.4.1 Role aware architecture of DMDSS

DMDSS supports role-aware menus. Every role has its own role-aware menu which enables the access only to its dedicated modules. Every DMSDSS user is granted one of the following roles: the data mining administrator, the business user and the developer. The last one was introduced for administrative and maintenance purposes. DMDSS allows the developer to maintain the catalogue of areas of analysis. The catalogue of areas of analysis is a group of database tables having the following advantages:

- The lists of areas of analysis are built dynamically, based on the current catalogue contents. This is used in the building of menus and lists of values.
- The name of the training set, attribute names and the name of classification attribute (only for classification) are stored in the database. This enables changes in data sources and its structure without changing of DMDSS program code.
- The translations of keywords used in models are stored in the database. One of the ODM API methods enables the access to the rules of the model and returns the rules as a string. Through the use of translating of the keywords (if, then, in, …) the model presentation can be adopted and changed without changing of DMDSS program code. In order to achieve higher flexibility, every area of analysis has its own keyword translations.

These advantages clearly reveal the flexibility of DMDSS for the introduction of new areas of analysis without changing the program code. The approach with the catalogue of the areas of analysis stored in the database ensures efficient maintenance process. In order to enable more flexible and environment-independent deployment, DMDSS also enables the developer to maintain usernames/passwords for ODM and other ODM parameters. All these parameters are also stored in the database.

The information support provided for the roles of data mining administrator and business user will be reflected in the following part of the chapter.

4.4.2 Classification

Classification-based areas of analysis were first supported by DMDSS. The classification method was chosen to be a test area for the concepts of GUI and the use of DMDSS and the first four development iterations were dedicated only to classification.

The example of the area of analysis for classification method is called "Customers classification". For the purpose of area of analysis customers are ranked into three categories: a good customer, an average customer and a bad customer. The aim of the area of analysis is to acquire the customer model for each customer category. This information enables business users to monitor characteristics of a particular customer category and plan better marketing campaigns for acquiring new customers. Within the DMDSS application additional areas of analysis for the purposes of mobile phone sales analysis, customer analysis and vendor analysis were also investigated.

The data mining administrator can create classification models by using model creation form (Figure 1). When creating the model they input a unique model name and a purpose of model creation. Beside that there are four algorithm parameters to be set before the model creation. The user can choose the value for each parameter from the interval which was defined as proper in the second stage of process model. At the bottom of the form there are recommended values for parameters to acquire a model with fewer or more rules: default settings for fewer rules in a model, and settings for more rules in a model. The examples of forms in the figures shown below are for the area of analysis called "Customers classification".

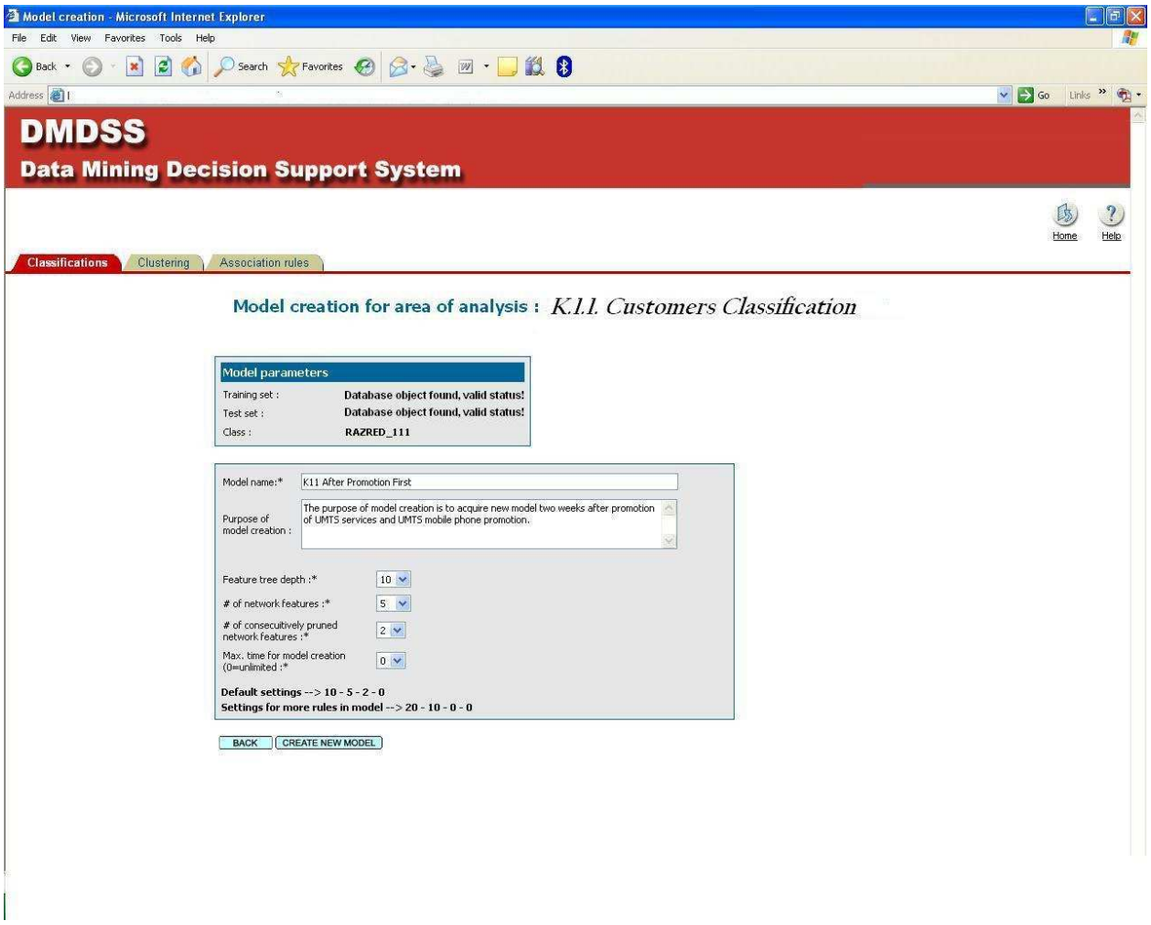


Fig. 1. A model creation form for classification

Model testing is performed automatically as the last phase of the model creation. Model testing is an evaluation process to perceive the quality of the model through using machine learning methods.

After the model creation, a data mining administrator can view and inspect the model. Model viewing is supported by two visualization techniques. The first technique is a table where classification rules are presented in a simple IF-THEN form. As already mentioned, keywords used in rules are translated in order to present the rules in a language more appropriate for the users. The second technique is decision trees, where classification rules are converted into decision trees showing equivalent information as rules. The decision trees technique is a graphical technique, which enables visual presentation of rules and for that reason it is very appropriate. While viewing and inspecting, the administrator can input comments for the model. As already mentioned, the role of the comments is to help the business users to understand and interpret the models better.

A data mining administrator can change the status of a model to a published status if the model quality reaches a certain level, and if the model is different from the previously created model of particular area of analysis. Business users can view only the models with published status.

Business users have access to a fewer functionalities than the data mining administrator. The form for model viewing for a business user (Figure 2) is slightly different from the form for model viewing for data mining administrator, but has similar general characteristics. Business users can also view rules in both visualization techniques as the data mining administrator. On the other hand, the form also enables access to some general information about the model: creation date, purpose of model creation, etc. The form also enables business users to view comments of a model written by the data mining administrator.

4.5 The experience of the use of DMDSS

DMDSS has now been in production for several months. During the first year of production there will be supervising and consultancy provided by the development team. Supervising and consultancy have the following goals:

- The role of data mining administrator will be supervised by the data mining consultant form development team, having expertise and experience in data mining. The employee responsible for that role has enough knowledge, but not enough experience yet. Supervising will mainly cover support at model evaluation and model interpretation for data mining administrator and business users;
- Support at defining and introducing new areas of analysis;
- Support at all stages of DMDSS process model before the introduction of new areas of analysis.

Business users use DMDSS at their daily work. They use patterns and rules identified in models as the new knowledge, which they use for analysis and decision process at their work. It is becoming apparent that they are getting used to DMDSS. According to their words they have already become aware of the advantages of continual use of data mining for analysis purposes. Based on the models acquired they have already prepared some changes in marketing approach and they are planning a special customer group focused campaign, based on the knowledge acquired in data mining models. The most important achievement after several months of usage is the fact that business users have really started to understand the potentials of data mining. Suddenly they have got many new ideas for

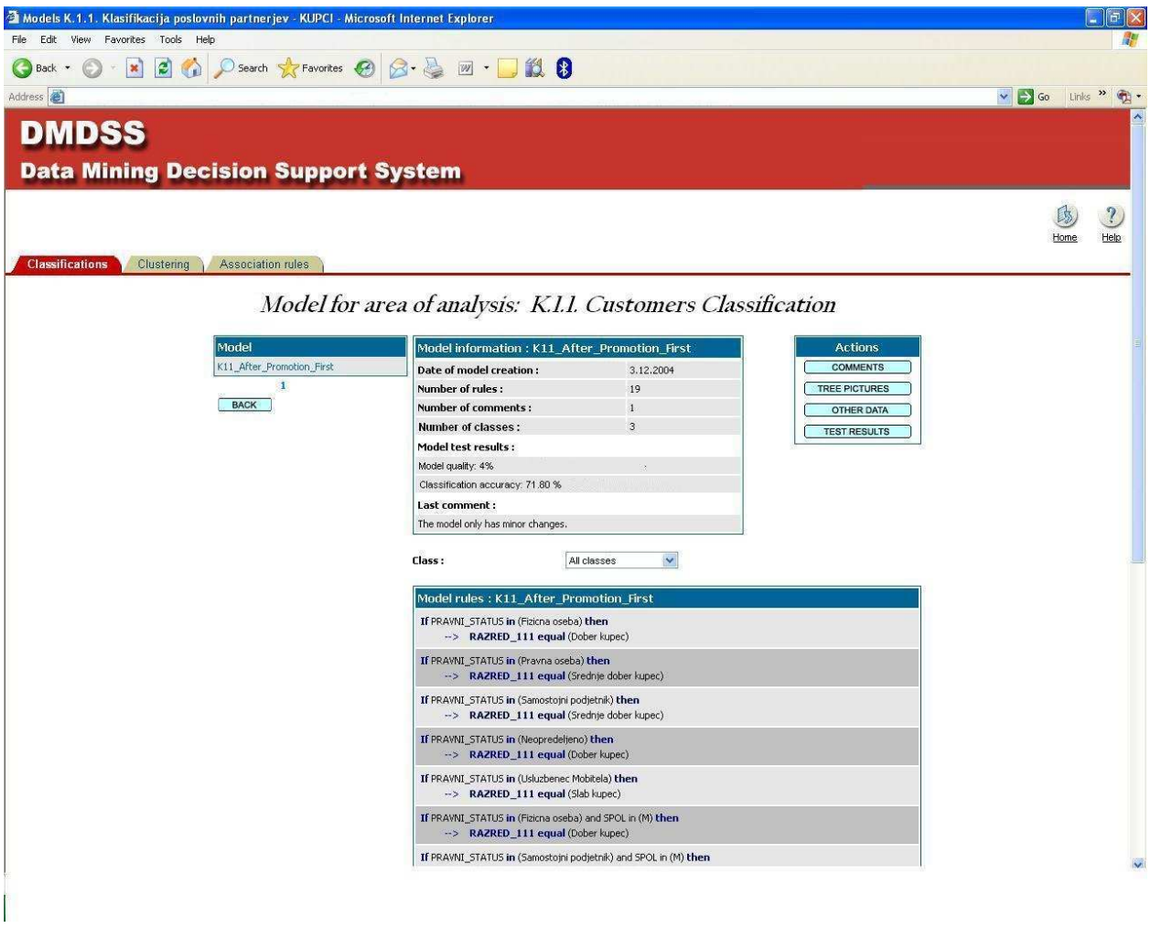


Fig. 2. A model viewing form for business users for classification

new areas of analysis, because they have started to realize how to define areas of analysis to acquire valuable results. The list of new areas of analysis will be made in several months, and after that it will be discussed and evaluated. Selected areas of analysis will then be implemented and introduced to DMDSS according to methodology introduced in the chapter.

The experience of the use of DMDSS has also revealed that business users need the possibility to make their own archive of classification rules. They also need to have an option to make their own comments to archived rules in order to record the ideas implied and gained by the rules. The future plan for classification model utilization is also to apply the model on new customers in order to predict the category a new customer potentially belongs to. These enhancements are planned to be implemented in the future.

4.6 Semantic contribution of the use of DMDSS

While designing and developing DMDSS and monitoring its use by the business users we have been considering and exploring the semantic contribution of the use of a data mining application system like DMDSS in a decision process and performing any kind of business analysis. For that reason one of our goals of the project was also to illustrate the semantic contribution of the use of DMDSS in decision processes. We decided to use the concept of data-model for that purpose.

A data-model is a concept which can be, among other things, used for describing a particular domain on a conceptual level (Bajec, 2001; Lavbic & Krisper, 2009; Sasa et al., 2008; Vavpotic et al., 2009). A meta-model shows domain concepts and relations between them. In this case the meta-model describes a decision process on the conceptual level with emphasis on demonstrating the contribution and the role of the use of DMDSS as data mining application system (Figure 3). UML class diagrams were used as technique for the meta-model. Decision support concepts are represented as classes and relations between them are represented as associations and aggregations. Concepts and relations, which in our opinion represent a contribution of the use of DMDSS in the decision process, are represented in a dotted line style.

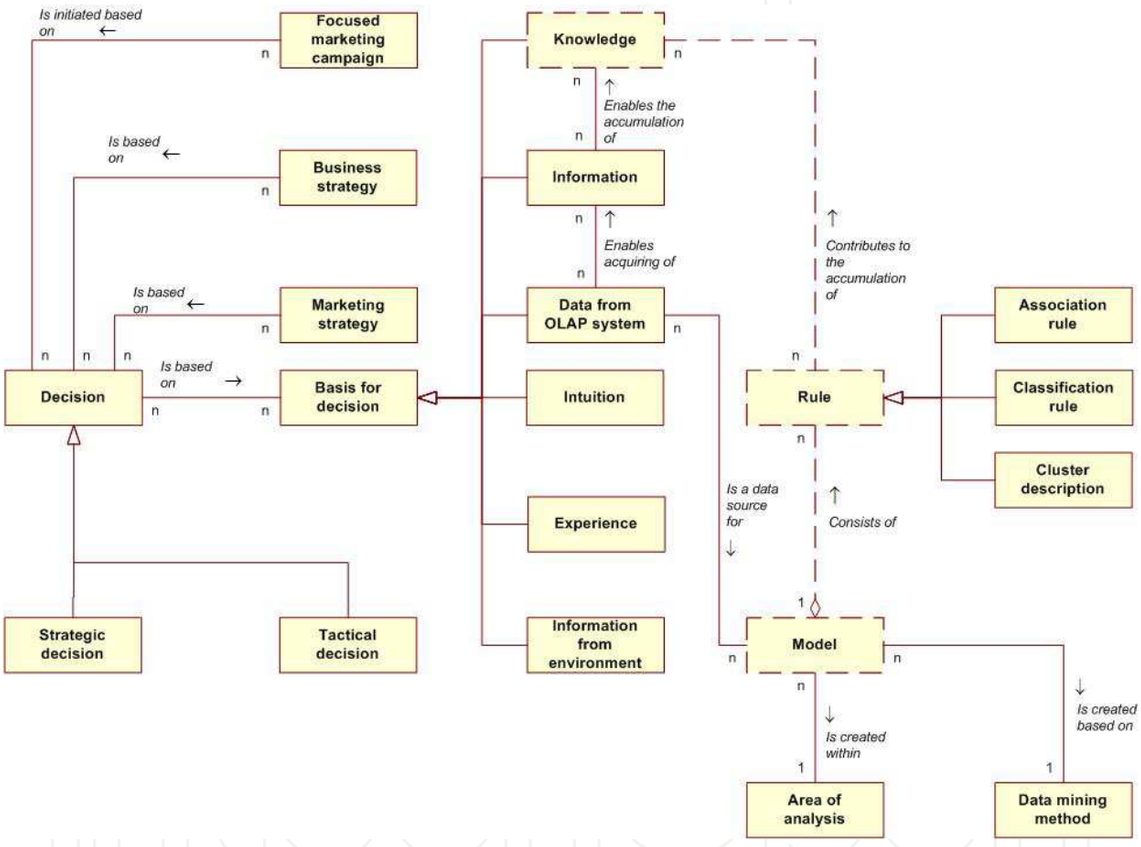


Fig. 3. A meta-model

The meta-model shows various concepts that influence the decision process and represent a basis for a decision. Information technology engineers often believe that decisions mostly depend on data from OLAP systems and other information acquired from information systems. It is true that they represent a very important basis for the decision, although in more than a few cases decisions mostly depend on factors like intuition and experience (Bohanec, 2001).

Knowledge is in our opinion probably the most important basis for the decision, because it enables the correct interpretation of data, i.e. acquiring of information. The contribution of the use of DMDSS and models and rules it creates is in contribution to the accumulation of

the knowledge acquired by models and their rules. A detailed description of decision process and creation of a detailed meta-model is beyond the scope of the chapter.

5. Summary and conclusions

DMDSS is a data mining application system which enables a decision support, based on the knowledge acquired from data mining models and their rules. The mission of DMDSS is to offer an easy-to-use tool which will enable business users to exploit data mining with only a basic level of understanding of the data mining concepts. DMDSS enables the integration of data mining into daily business processes and decision processes through supporting several areas of analyses.

The experience of the use of DMDSS has revealed that "traditional" data mining expert role is different, according to the data mining software tool approach. A DMDSDS process model divides the traditional data mining expert role into a data mining administrator role and a data mining consultant. The data mining consultant provides support at defining and introducing of new areas of analysis. The data mining administrator executes daily model creation and provides support for business users. The former must have expertise in data mining; the latter must have enough knowledge of data mining to evaluate models acquired and detect problems at the model creation.

The experience of the use of DMDSS has also revealed that it has become a tool regularly used by business users at decision process and performing various kinds of analyses. They are getting used to DMDSS and they have become aware of the advantages of the continual use of data mining for analysis purposes. After several months of usage, business users have started to realize how to define the areas of analysis to acquire valuable results. We believe that we have succeeded in achieving the optimal data mining process organization and infusion of data mining into decision processes with DMDSS.

The first results of the use of DMDSS are some changes planned in the marketing approach and a special customer group focused campaign based on the knowledge acquired in data mining models. Another result is the revealing of bad data quality, which is a typical side-effect result for the use of data mining. For some areas of analysis bad data quality has been detected in the development of DMDSS and measures at the sources of data have been taken to improve data quality.

Although DMDSS is a rather new application system, there exists a plan for future development of DMDSS. On one hand, there is a list of new areas of analysis being built up by business users, on the other hand there are also enhancements planned in the area of functionalities of DMDSS. There are several directions we intend to explore in the future. We intensively follow the development of the application platform of choice, Oracle Data Mining and accompanying tools, which have already gained certain level of maturity. We intend to provide our users with more data mining methods (e.g. decision trees, rules, …) when they become available. We believe that both satisfying the user requirements as well as providing them with a choice of new data mining methods will contribute to better results of the use of DMDSS.

6. References

Aggarwal, C.C. (2002). Towards-Effective and Interpretable Data Mining by Visual Interaction, *SIGKDD Explorations*, 3, 11-22

Bajec, M. & Krisper, M. (2005). A Methodology and Tool Support for Managing Business Rules in Organisations, *Information Systems*, 30, 423-443

Bayardo, R. & Gehrke, J.E. (2001). Report on the Workshop on Research Issues in Data Mining and Knowledge Discovery Workshop (DMKD 2001), *SIGKDD Explorations*, 3, 43-44

Bohanec, M. (2001). What is Decision Support?. *Proceedings Information Society IS-2001: Data Mining and Decision Support in Action!* (pp. 86-89), Ljubljana, Slovenia

Bose, R. & Sugumaran, V. (1999). Application of Intelligent Agent Technology for Managerial Data Analysis and Mining, *The DATABASE for Advances in Information Systems*, 30, 77-94

Chen, M.S., Han, J. & Yu, P.S. (1996). Data Mining: An Overview from a Database prespective, *IEEE Transactions on Knowledge and Data Engineering*

Chu, R. (2003). XML for Analysis. *KDD-2003 Workshop on Data Mining Standards, Services and Platforms (DM-SSP 03), http://www.ncdm.uic.edu/workshops/dm-ssp03.htm*, Washington DC, USA: ACM

Clifton, C. & Thuraisingham (2001). Emerging Standards for Data Mining, *Computer Standards & Interfaces*, 23, 187-193

Farnstrom, F., Lewis, J. & Elkan, C. (2000). Scalability for Clustering Algorithms Revisited, *SIGKDD Explorations*, 2, 51-57

Fayyad, U.M. & Uthurusamy, R. (1996). Data Mining Knowledge Discovery in Databases (editorial), *Communications of the ACM*, 39, 24-26

Fayyad, U. & Uthurusamy, R. (2002). Evolving Data Mining into Solutions for Insight, *Communications of ACM*, 45, 28-31

Fayyad, U., Haussler, D. & Stolorz, P. (1996). Mining Scientific Data, *Communications of ACM*, 39, 51-57

Furlan, S. & Bajec, M. (2008). Holistic approach to fraud management in health insurance, *Journal of Information and Organizational Sciences*, 32, 99-114

Geist, I. (2002). A Framework for Data Mining and KDD. *Proceedings ACM Symposium on Applied Computing* (pp. 508-513), Madrid, Spain: ACM

Glaymour, C., Madigan, D., Pregibon, D. & Smyth, P. (1996). Statistical Inferrence and Data Mining, *Communications of the ACM,* 39, 35-41

Goebel, M. & Gruenwald, L. (1999). A Survey of Data Mining Knowledge Discovery Software Tools, *SIGKDD Explorations*, 1, 20-33

Grossman, R. (2003). KDD-2003 Workshop on Data Mining Standards, Services and Platforms (DM-SSP 03), *SIGKDD Explorations*, 5, 197

Grossman, L.G., Hornick, M.F. & Meyer, G. (2002). Data Mining Standard Initiatives, *Communications of ACM*, 45, 59-61

Han, J., Altman, R.B., Kumar, V., Mannila, H. & Pregibon, D. (2002). Emerging Scientific Applications in Data Mining, *Communications of ACM*, 45, 54-58

Hasti, T. & Tibisharani, R. (2001). The Elements of Statistical Learning, Springer

Heinrichs, J. & Lim, J.S. (2003). Integrating Web-based Data Mining Tools with Business Models for Knowledge Management, *Decision Support Systems*, 35, 103-112

Hirji, K.K. (2001). Exploring Data Mining Implementation, *Communications of ACM*, 44, 87-93

Hofmann, T. & Buhmann, J. (1997). Active Data Clustering. *Proceedings Advances in Neural Information Processing Systems* (pp. 528-534), Denver, USA

Holsheimer, M. (1999). Data Mining by Business Users: Integrating Data Mining in Business Process. *Proceedings International Conference on Knowledge Discovery and Data Mining KDD-99* (pp. 266-291), San Diego, USA: ACM

Holsheimer, M., Kersten, M. & Toivonen, H. (1995). A Perspective on Databases and Data Mining. *Proceedings International Conference on Knowledge Discovery and Data Mining KDD-1995*, Montreal, Canada: ACM

Hornick, M. (2003). Java™ Data Mining (JSR-73): Overview and Status. *KDD-2003 Workshop on Data Mining Standards, Services and Platforms (DM-SSP 03), http://www.ncdm.uic.edu/workshops/dm-ssp03.htm*, Washington DC, USA: ACM

JSR-73 Expert Group (2004). *Java™ Specification Request 73: Java™ Data Mining (JDM)*, Oracle Corporation and Sun Microsystems, Inc.

Kaufmann, L. & Rousseeuw, P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*, New York: John Wiley & Sons

Kohavi, R. & Sahami, M. (2000). KDD-99 Panel Report: Data Mining into Vertical Solutions, *SIGKDD Explorations*, 1, 55-58

Kohavi, R., Rothleder, N.J. & Simoudis, E. (2002). Emerging Trends in Business Analytics, *Communications of ACM*, 45, 45-48

Kononenko, I. (2001). Machine Learning for Medical Diagnosis: History, State of the Art and Perspective, *Artificial Intelligence in Medicine,* 23, 89-109

Kukar, M. (2003). Transductive Reliability Estimation for Medical Diagnosis, *Artificial Intelligence in Medicine,* 29, 81-106

Kukar, M., Kononenko, I., Groselj, C., Kralj, K. & Fettich, J. (1999). Analysing and Improving the Diagnosis of Ischaemic Heart Disease with Machine Learning, *Artificial Intelligence in Medicine,* 16, 25-50

Kumar, A. & Kantardzic, M. (2003). Grid Application Protocols and Services for Distributed Data Mining. *KDD-2003 Workshop on Data Mining Standards, Services and Platforms (DM-SSP 03), http://www.ncdm.uic.edu/workshops/dm-ssp03.htm*, Washington DC, USA: ACM

Lee, W. & Stolfo, S. (1998). Data Mining Approcahes for Intrusion Detection. *Proceedings USENIX Security Symposium*, San Antonio, USA

Lavbic, D. & Krisper, M. (2009). Rapid Ontology Development. *Proceedings of the 19th European-Japanese Conference on Information Modeling and Knowledge Bases*, Maribor, Slovenia

Li, T., Li, Q., Zhu., S. & Ogihara, M. (2002). A Survey on Wavelet Applications in Data Mining, *SIGKDD Explorations*, 4, 49-68

Ma, Y., Liu, B., Wong, C.K., Yu, P.S. & Lee, S.M. (2000). Targeting the Right Students Using Data Mining. *Proceedings International Conference on Knowledge Discovery and Data Mining KDD-2000* (pp. 457-464), Boston, USA: ACM

Meyer, G. (2003). PMML Version 3 – Overview and Status. *KDD-2003 Workshop on Data Mining Standards, Services and Platforms (DM-SSP 03), http://www.ncdm.uic.edu/workshops/dm-ssp03.htm*, Washington DC, USA: ACM

Ng, K. & Liu, H. (2000). Customer Retention via Data Mining, *Artificial Intelligence Review*, 14, 569-590

Sasa, A., Juric, M. & Krisper, H. (2008). Service Oriented Framework for Human Task Support and Automation, *IEEE Transactions on Industrial Informatics*, 4, 292-302

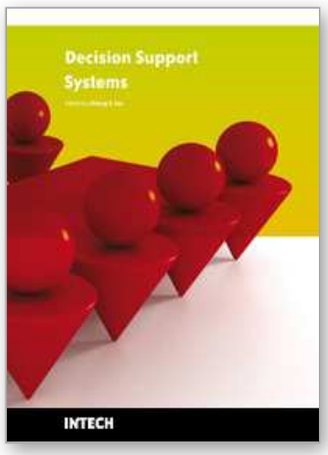
Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining, *Journal of Data Warehousing*, 5, 13-22

Srivastava, J., Cooley, R., Deshpande, M. & Tan, P.N. (2000). Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, *SIGKDD Explorations*, 1, 12-23

Vavpotic, D. & Bajec, M. (2009). An approach for concurrent evaluation of technical and social aspects of software development methodologies, *Information and Software Technology*, 51, 528-545

Witten, I.H. & Frank, E. (2000). *Data Mining*, Morgan-Kaufmann, 978-1558605527, New York

Yarmus, J.S. (2002). *ABN: A Fast, Greedy Bayesian Network Classifier*, Oracle White Paper, Burlington, MA: Oracle Coorporation



Decision Support Systems

Edited by Chiang S. Jao

ISBN 978-953-7619-64-0
Hard cover, 406 pages
Publisher InTech
Published online 01, January, 2010
Published in print edition January, 2010

Decision support systems (DSS) have evolved over the past four decades from theoretical concepts into real world computerized applications. DSS architecture contains three key components: knowledge base, computerized model, and user interface. DSS simulate cognitive decision-making functions of humans based on artificial intelligence methodologies (including expert systems, data mining, machine learning, connectionism, logistical reasoning, etc.) in order to perform decision support functions. The applications of DSS cover many domains, ranging from aviation monitoring, transportation safety, clinical diagnosis, weather forecast, business management to internet search strategy. By combining knowledge bases with inference rules, DSS are able to provide suggestions to end users to improve decisions and outcomes. This book is written as a textbook so that it can be used in formal courses examining decision support systems. It may be used by both undergraduate and graduate students from diverse computer-related fields. It will also be of value to established professionals as a text for self-study or for reference.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen